

Quantum Mechanics in Terms of Realism

Arthur Jabs

Abstract

We expound an alternative to the Copenhagen interpretation of the formalism of nonrelativistic quantum mechanics. The basic difference is that the new interpretation is formulated in the language of epistemological realism. It involves a change in some basic physical concepts. The ψ function is no longer interpreted as a probability amplitude of the observed behavior of an elementary particle but as an objective physical field representing the particle itself. The particles are thus extended objects whose extension varies in time according to the variation of ψ . They are considered as fundamental regions of space with no internal structure. This implies some kind of nonlocality. Symmetrization of the configuration space wave function is interpreted as a mathematical description of a physical process, which also leads to nonlocal effects. Special consideration is given to the problem of measurement, the reduction process, Schrödinger's cat, Wigner's friend, the Einstein-Podolsky-Rosen correlations, field quantization and quantum-statistical distributions. Experiments to distinguish the proposed interpretation from the Copenhagen interpretation are pointed out.

Key words: foundations of quantum mechanics, interpretation, epistemological realism, particles as extended objects, wave function symmetrization, measurement problem, reduction of the wave function, nonlocality, Einstein-Podolsky-Rosen problem, Bell inequality

CONTENTS

1. INTRODUCTION: GENERAL PRINCIPLES	
1.1 Difficulties in Present-Day Quantum Theory	3
1.2 Realism	5
1.3 Wavepackets	8
2. SPECIFIC POSTULATES: SINGLE PARTICLES	
2.1 The Double-Slit Experiment	12
2.2 The Action Probability	14
2.3 The Internal Structurelessness of Wavepackets	18
2.4 Normalization and Quantization	23
2.5 The Heisenberg Relations	25
3. SPECIFIC POSTULATES: SYSTEMS OF PARTICLES	
3.1 Similar Wavepackets: Coalescence and Splitting	28
3.2 Condensed Wavepackets	33
3.3 Dissimilar Wavepackets: Aggregates	35
3.4 Summary of Principles and Postulates	39
4. MEASUREMENTS AND TRANSITIONS	
4.1 The Copenhagen Measurement Postulates	41
4.2 Derivation of $ \psi ^2 d^3x$ as an Action Probability	44
4.3 Continuation: The Time Variable	47
4.4 The Stern-Gerlach Experiment	49
4.5 Criticism of von Neumann's Reduction Axiom	53
4.6 The Underlying Physical Problem	55
4.7 Determinism and Indeterminism	56
5. THE EINSTEIN-PODOLSKY-ROSEN PROBLEM	
5.1 The EPR Problem and Nonlocality	59
5.2 Relativity and Causality	63
5.3 Bell's Inequality and Experiments	65
5.4 Crucial Experiments	70
6. QUANTUM STATISTICS WITH WAVEPACKETS	
6.1 Field Quantization	74
6.2 The Many Aspects of the Condensed Wavepackets	76
6.3 The Balance Relation	79
6.4 The Bose and Fermi Distributions	83
6.5 Quantum Count Fluctuations	86
Acknowledgement	91
APPENDIX A: COLLECTION OF WAVEPACKET SPREADING FORMULAS	92
APPENDIX B: PROOF OF THE BELL INEQUALITY	99
APPENDIX C: EPR JOINT PROBABILITY FORMULAS	101
APPENDIX D: EPR PROBABILITIES IN DIFFERENT SYSTEMS OF EIGENFUNCTIONS	105
References	109

Sage, was du willst, solange dich das nicht verhindert, zu sehen, wie es sich verhält. (Und wenn du das siehst, wirst du Manches nicht sagen.)

Ludwig Wittgenstein

1. INTRODUCTION: GENERAL PRINCIPLES

1.1 Difficulties in Present-Day Quantum Theory

There is no doubt that quantum theory is one of the most successful physical theories. Yet there is also no doubt that it contains serious difficulties. These difficulties are nowadays felt more and more strongly by those concerned with the unification of quantum theory and relativity and the future basis of physics. The difficulties may be divided into two kinds: conceptual and mathematical.

The conceptual difficulties are related to the so-called Copenhagen interpretation. Any physical theory consists of a mathematical formalism, that is, a set of mathematical symbols and the rules for connecting these among themselves, and a set of interpretation rules, connecting the symbols of the mathematical formalism with the elements of our sensuous experience. The Copenhagen interpretation represents that set of interpretation rules that has been generally accepted and is presented more or less explicitly in the present textbooks on quantum mechanics. Actually, it is difficult to say who exactly constitutes the “Copenhagen school” supporting the Copenhagen interpretation: certainly Bohr and Heisenberg, but also Dirac, Pauli and von Neumann.^(1,2) Also, many versions of “the Copenhagen interpretation” are met, from orthodox to liberal, when different authors or textbooks are consulted. As a rule we shall always consider the most orthodox version, which is always the most consistent one.

The difficulties of the Copenhagen interpretation may be characterized in the following way.

1) The wave function $\psi(\mathbf{x}, t)$ is not taken as an objective physical field like the electromagnetic field, but as a probability amplitude. And the probabilities to which it refers are not the probabilities that something is true or something will happen, whether it is observed or not (as in statistical mechanics), but are the probabilities of specified outcomes of measurements or observations. Moreover, and most important, the observer is not just another physical object, but is the linguistic ego, something which nowhere appears as a symbol in the formalism. An electron does not have an exact location as long as we do not observe it, but it does have one when we observe it. “The ‘trajectory’ arises only by our observing it” (“Die ‘Bahn’ entsteht erst dadurch, daß wir sie beobachten”⁽³⁾). In this way the Copenhagen interpretation speaks of position, and of momentum, angular momentum etc. only as “observables”, not as real properties, which the objects have independent of whether or not we observe them. Therefore the observer and the measurement are indispensable elements in defining the theory. In all other physical theories the observer’s only function is to test and apply the theory, not to define it. The Copenhagen interpretation thus rejects the language

of epistemological realism. In my opinion, this is the most serious difficulty of that interpretation.

2) The Copenhagen interpretation maintains that it is impossible, in principle, to explain the probabilistic behavior in quantum physics by some underlying deterministic processes and variables that would specify the physical situations in more detail than it is done by the ψ function. This means that the Copenhagen interpretation rejects determinism. This point will be examined further in Sec. 4.7.

3) In the Copenhagen view concepts that refer to point particles (single, sharp position, etc.) are applied to the microscopic objects, whereas the ψ function is the solution of a field equation, namely a partial differential equation, like Schrödinger's, and may show up wave-like behavior. This has been called the "wave-particle duality", and it has been asserted that a unifying picture of the microscopic objects cannot exist.

As is well known Einstein, von Laue, Schrödinger, Planck, and de Broglie never accepted the Copenhagen interpretation. In fact, although the Copenhagen interpretation seems to be without apparent logical inconsistencies any really thoughtful scholar finds it difficult to digest. Inquiring students are usually silenced by authoritarian statements such as that they do not yet know enough and will understand later, or that their questions are not relevant to physics.

In particular, the defenders of the Copenhagen interpretation generally assert that any unifying picture is really unnecessary. They say that the formalism plus some working rules for its application give us the correct prescriptions for calculating the probabilities of the outcomes of any experiment, and that is all we want. But as if some tectonic tensions were felt discussions continually arise regarding the foundations of quantum mechanics. Moreover, one always notices that in the course of historical development as well as in the mind of any particular scientist the final mathematical formalism describing a set of physical phenomena emerges from a more or less pictorial view, conception or model. A good picture is very helpful since it has the same logical structure as the region of reality which it represents, and it leads to a correct mathematical formulation of this reality. An example is Faraday's intuitive picture of lines of force and their subsequent mathematical formulation by Maxwell. "It is mainly with the hope of making these [Faraday's] ideas the basis of a mathematical method that I have undertaken this treatise", writes Maxwell.⁽⁴⁾ A bad picture leads to no or only to a partially correct formalism. In this latter case it may happen that the emerging formalism in its initial stage describes the known phenomena correctly, but when it is developed further to include more and more experimental facts it sooner or later comes off the track. This is what I think has happened to quantum theory. I think that the lack of a good picture is responsible for the mathematical difficulties, and that their solution will emerge only from a solution of the conceptual difficulties.

The mathematical difficulties of present-day quantum theory arise with the attempt to extend nonrelativistic quantum mechanics into the relativistic domain, that is, into quantum electrodynamics and relativistic quantum field theory. Here, divergent integrals have shown up in the perturbation expansions as solutions

of the basic equations. Even if these integrals are made finite by means of the renormalization procedures or are avoided by means of Epstein-Glaser methods,⁽⁵⁾ nobody knows whether the expansions converge, and nobody has found an exact solution of the equations including interactions in the real world of 3+1 dimensions, although enormous efforts have been undertaken.⁽⁶⁻⁸⁾ This raises the suspicion that the solution will lie in an entirely different direction. Thus Dirac writes:

I feel pretty sure that the changes which will be needed to get over the present difficulties facing quantum theory and appearing as a resistance between the quantum theory and relativity will be very drastic just as drastic as the change from Bohr orbits to the quantum mechanics of Heisenberg and Schrödinger and therefore one should not become too much attached to the present quantum mechanics. One shouldn't build up one's whole philosophy as though this present quantum mechanics were the last word. If one does that, one is on very uncertain ground and one will in some future time have to change one's standpoint entirely.⁽⁹⁾

1.2 Realism

While Dirac says nothing about the nature of the expected changes in quantum mechanics, Einstein's critique⁽¹⁰⁻¹⁶⁾ is more specific. Einstein writes:

But in any case my conception starts from a thesis which is strongly rejected by most present-day theoreticians: *There is something like the "real state"* of a physical system, which independent of any observation or measurement exists objectively and which can in principle be described by means of physical terms [Which adequate terms or basic concepts have to be employed for this is in my opinion unknown at the present moment (material points? field? concepts that have still to be invented?)]. Because of its "metaphysical" nature, this thesis of reality does not have the purpose of providing a statement of fact: it has really only a *programmatic* character. However, everybody, including the quantum theoreticians, sticks consistently to this thesis of reality so long as he does not discuss the foundations of quantum theory. Nobody doubts, for example, that there has been at a certain time a certain position of the moon's center of gravity even if no real or potential observer existed.⁽¹⁵⁾

This thesis of Einstein's is what we mean by epistemological realism. We do not attempt to give a fool-proof definition of realism. We emphasize, however, that the type of realism adopted here does not mean that the physical objects and their properties exist independent of whether or not we observe them (this would be naive realism) but only means that the laws of nature can be formulated as if that were the case. Realism, as it is meant here, is not considered as a matter to be proved or disproved, it is a way of speaking, a language. It is in fact the manner in which our language works in normal (not philosophical) use, in the sense of Wittgenstein:

For *this* is what disputes between Idealists, Solipsists and Realists look like. The one party attacks the normal form of expression as if they were attacking

a statement, the others defend it, as if they were stating facts recognized by every reasonable human being.⁽¹⁷⁾

If we were to distinguish this type of realism from naive realism we would call it linguistic realism or enlightened realism, in the sense that historians speak of enlightened monarchism.

Von Laue,⁽¹⁸⁾ Schrödinger,^(19–21) and Planck^(22,23) have always shared a realistic attitude with Einstein. And in the course of years the number of physicists who openly advocate realism in quantum theory has continually increased. Jammer's book⁽²⁴⁾ already quotes Bohm, Bunge, de Broglie, Jaynes, Ludwig, Popper and Renninger. And we want to add the papers by Janossy,⁽²³⁾ de Broglie,⁽²⁵⁾ Bunge and Kalnay,^(26–28) Bell,^(29–32) Rayski,⁽³³⁾ Lévy-Leblond,⁽³⁴⁾ Stapp,⁽³⁵⁾ Roberts,⁽³⁶⁾ Maxwell,⁽³⁷⁾ Burgos,⁽³⁸⁾ Popper,⁽³⁹⁾ Pearle,⁽⁴⁰⁾ Bohm, Hiley and Kaloyerou,⁽⁴¹⁾ Rohrlich,⁽⁴²⁾ Dorling⁽⁴³⁾ and Dieks.^(44–46) Actually, it is difficult to do justice to everybody because there are several types of realism, because there is a continuous variety from very outspoken to rather casual statements in favor of realism in the various papers, and because the problem of a realistic interpretation is often mixed up with a change in the formalism to be interpreted. Scientists whom I found particularly outspoken are Popper, Bunge and Bell. Bell⁽³¹⁾ in particular postulated “*beables*” to replace the “*observables*”, and his work will concern us in Sec. 5 when we discuss the Einstein-Podolsky-Rosen (EPR) problem. In fact it is mainly through the work of Bohm and Bell on the EPR problem and the relating series of experiments carried out since 1972 that the question of realism has received new interest in recent times.

Thus in the present work we propose to overcome the conceptual difficulties of quantum mechanics by interpreting the formalism in terms of realism. We replace the interpretation rules of the Copenhagen interpretation by interpretation rules that are formulated in the language of epistemological realism. What is interpreted is the same as what is interpreted by the Copenhagen school, namely the standard formalism of nonrelativistic quantum mechanics, as for example presented in Messiah's textbook.⁽⁴⁷⁾ In other words, we separate the problem of changing the interpretation from the problem of changing the formalism, and we restrict ourselves to the interpretation problem. The treatment of the mathematical problems is beyond the scope of this work.

Of course it is to be expected that ultimately both the formalism and the interpretation will have to be changed, and it may be legitimately asked whether the restriction to a change in the interpretation only makes sense. It certainly cannot solve all problems; nevertheless it can solve a number of them, as I intend to show in this essay. Anyway, changing both the interpretation and the formalism at one time to me appears too big a task. There exist a number of proposals in that direction (e.g. Jánossy,⁽²³⁾ Bohm,⁽⁴⁸⁾ and Pearle⁽⁴⁰⁾) but in my opinion they do not go far enough, they are “not crazy enough” in Bohr's words.⁽⁴⁹⁾ Thus this work is an attempt to see how far one can get with a novel interpretation without proposing any new formulas. Nevertheless, since the prescriptions of which mathematical formula is to be used in which physical situation are different in the two interpre-

tations it will be possible to point out a number of experiments where the realistic and the Copenhagen quantum mechanics predict different results (Sec. 5.4).

Our main result is the demonstration that a realistic interpretation is possible, in contradiction to what is asserted by the Copenhagen school. Other interpretations of the unmodified standard formalism in terms of realism might perhaps also be possible, differing from the present one in the details of the elaboration, although to my knowledge there is none that has been elaborated to a comparable extent.

It is mainly for convenience of presentation that the nonrelativistic formalism is chosen, with the familiar Schrödinger equation and wave function as the basis of the reinterpretation. We think that the concepts developed will prove fruitful in the relativistic domain as well; at least we do not know of any argument that would point to the contrary. Everything that can be described by the Schrödinger equation can also be described by the Klein-Gordon, Dirac etc. equation. Thus we might as well have used the Klein-Gordon, Dirac, etc. equation, and used a Lorentz scalar, spinor, vector etc. instead of the Schrödinger scalar $\psi(\mathbf{x}, t)$.

Now, the first step in our realistic program is to take the function $\psi(\mathbf{x}, t)$ (in the Schrödinger picture, say) as an objective physical field comparable in this respect to the function $F_{\mu\nu}(\mathbf{x}, t)$ as the electromagnetic field.

However, not every ψ function is so taken, but only that which refers to an elementary particle. And this ψ function is no longer the probability amplitude of the observed behavior of the particle but is the individual elementary particle itself. Elementary particles are defined here as those that are listed in the Tables of Particle Properties.⁽⁵⁰⁾ Of course, this definition is based more on general consensus than on clearly defined principles. Essentially, what we want to say is that there are two kinds of objects, “fundamental” objects and others, and that our notion of ψ refers only to the fundamental objects. For the time being we adopt the conventional definition in order to be definite.

This implies (1) that ψ does not represent nuclei, atoms, molecules, centers-of-mass, nor quarks, as it does in the Copenhagen interpretation, (2) that ψ is not merely a calculational device for calculating the probabilities of specified outcomes of observations, (3) that it does not merely describe “knowledge”,⁽⁵¹⁾ and (4) that it does not describe an ensemble of particles, as in the statistical interpretation.⁽⁵²⁾

Thus there are no longer two different elements in our interpretation, namely, the particle and the field associated with it; rather these two are the same. There is no wave-particle duality. The function $\psi(\mathbf{x}, t)$ describes the particle like a region of excessive water density $\rho(\mathbf{x}, t)$ in the sky describes a cloud. And the particle has its properties (“beables”) even when they are not observed. Which properties these are is another question, and in fact we shall see that the new properties are different from the old ones.

Thus, for example, single sharp values of position or momentum are no longer among the properties. The properties that we shall have to attribute to the elementary particles in order to carry through the realistic program, and the concepts to be introduced in the subsequent sections at first sight might appear rather strange.

This is the price one has to pay. Nowadays we are perhaps willing to pay more than the physicists in Schrödinger's time when the successes rather than the difficulties of quantum mechanics stood in the foreground. Also, it is not to be expected that the difficulties that have beset quantum theory for more than 60 years could be overcome by some cheap trick.

On the other hand, we try to show that the price is not too high in that once the new postulates are accepted a coherent picture emerges. The conceptual mist dissolves and the view opens for a fresh outlook. Quantum mechanics is formulated in the same language as classical mechanics and any other physical theory, and the power that lies in the realistic language is available for quantum mechanics too.

1.3 Wavepackets

The identification of an elementary particle with a field means that quantum mechanics becomes a field theory. The Schrödinger equation, or any of the quantum equations of motion, in any case is a field equation, that is, a *partial* differential equation, with the solution ψ depending on the four independent variables x, y, z and t . The equations of motion of the point particles of classical mechanics, on the contrary, are *ordinary* differential equations for the three functions $x(t), y(t)$ and $z(t)$. On this point Einstein writes:

The most difficult point for such a field theory at present is how to include the atomic structure of matter and energy. For the theory in its basic principles is not an atomic one in so far as it operates exclusively with continuous functions of space, in contrast to classical mechanics whose most important feature, the material point, squares with the atomistic structure of matter.

The modern quantum theory, as associated with the names of de Broglie, Schrödinger, and Dirac, which of course operates with continuous functions, has overcome this difficulty by means of a daring interpretation, first given in a clear form by Max Born: - the space functions which appear in the equations make no claim to be a mathematical model of atomic objects. These functions are only supposed to determine in a mathematical way the probabilities of encountering those objects in a particular place or in a particular state of motion, if we make a measurement. This conception is logically unexceptionable, and has led to important successes. But unfortunately it forces us to employ a continuum of which the number of dimensions is not that of previous physics, namely 4, but which has dimensions increasing without limit as the number of the particles constituting the system under examination increases. I cannot help confessing that I myself accord to this interpretation no more than a transitory significance. I still believe in the possibility of giving a model of reality, a theory, that is to say, which shall represent events themselves and not merely the probability of their occurrence. On the other hand, it seems to me certain that we have to give up the notion of an absolute localization of the particles in a theoretical model. This seems to me to be the correct theoretical interpretation of Heisenberg's indeterminacy relation. And yet a theory may perfectly well exist, which is in a genuine sense an atomistic one (and not merely on the basis of a particular

interpretation), in which there is no localizing of the particles in a mathematical model. For example, in order to include the atomistic character of electricity, the field equations only need to involve that a three-dimensional volume of space on whose boundary the electrical density vanishes everywhere, contains a total electrical charge of an integral amount. Thus in a continuum theory, the atomistic character could be satisfactorily expressed by integral propositions without localizing the particles which constitute the atomistic system.

Only if this sort of representation of the atomistic structure be obtained could I regard the quantum problem within the framework of a continuum theory as solved.⁽⁵³⁾

The idea that elementary particles are extended objects has repeatedly appeared in the literature. However, the size has always been considered to be fixed, for example equal to the Compton wavelength $\lambda_C = \hbar/mc$, the classical (electron) radius $r_{cl} = e^2/mc^2$, or the Planck length $l_P = (\hbar G/c^3)^{1/2}$. In the interpretation presented here the size of any individual particle is variable, namely equal to the size of the ψ function traditionally associated with it. Mathematically, the ψ function need not have a sharp boundary but for our purposes it may be considered to have the extension given by the usual standard deviation $\Delta x := \langle (x - \langle x \rangle)^2 \rangle^{1/2}$, which varies in time according to the variation of $\psi(\mathbf{x}, t)$.

In order to emphasize that we consider the elementary particles not as pointlike but as extended objects we call them “wavepackets,” which is now a technical term of our interpretation. The function $\psi(\mathbf{x}, t)$ is the mathematical description of an elementary wavepacket. Sometimes we shall neglect the difference between the wavepacket and its mathematical description and just call ψ a wavepacket. The term “wavepacket” does not, however, in any way mean a restriction to a linear superposition of plane waves, and even when it is mathematically expressed as such it does not mean that plane waves are physical constituents of the particles. It just means the region(s) of non-vanishing ψ . A particle thus may be regarded as a “matter pulse”, on an equal footing with an electromagnetic radiation pulse. Indeed, in our interpretation the radiation pulse, under certain conditions, *is* also a particle, namely a photon. For each kind of particle a special type of “matter” field is conceived: there is an electron field, a photon field, a proton field, etc., as is familiar from quantum field theory.

However, our concept of a wavepacket goes further; there are wavepackets in our interpretation that are not elementary particles. In Secs. 3.1 and 3.2 below we shall consider systems of similar (identical) elementary particles and the process of condensation of these into one single wavepacket.

At the time when Schrödinger developed his wave mechanics he already tried to identify the electron with a wavepacket of small but finite extension constructed by superposition of particularly simple solutions of the Schrödinger equation.⁽⁵⁴⁾ The physicists of his time did not become convinced of this idea because in general the wavepackets spread out in time and for other reasons. Though Schrödinger was aware of the difficulties⁽⁵⁵⁾ and although he admitted not knowing how to

overcome them he never gave up the belief that a realistic wavepacket picture should be possible.^(56–58) What makes me so bold, that I bring up the idea of a realistic wavepacket again? It is the fact that I am willing to pay a higher price and to accept a rather drastic change in the basic concepts.

Let us consider some popular objections against the general idea that wavepackets can be models of elementary particles.

The first objection is that ψ can take on complex values and hence cannot represent a physical object. The objection is very weak and has been dealt with conclusively by Bunge⁽⁵⁹⁾: it dissolves as soon as one realizes that a complex function is nothing but a couple of real functions united in a convenient way, and as soon as one remembers other situations in physics where complex quantities are used; for example the impedance of an electric circuit $Z = R + iX$. Notice that we never directly measure “the electron” but only its charge, velocity etc., namely its properties. Thus “the electron” may be represented by a complex function in the mathematical formalism provided only that the values of its directly measurable properties are represented by real numbers. This is related to the old philosophical distinction between substance and attributes.

The second objection is that a free wavepacket of de Broglie waves inevitably spreads out in time. A corresponding difficulty already exists in ordinary quantum mechanics where the wavepacket gives the probability that the position of the particle will show up. If, for example, the wavepacket of an electron which moves perpendicularly through successive registering screens were to spread out rapidly, the “positions” of the electron on the successive screens would be distributed erratically in a rapidly widening cone, and would not exhibit anything like a straight track. The actually observed straight track is usually ascribed to the facts that (1) the spreading of the wavepacket numerically is negligibly small because the packet need not have a smaller extension than that of an observed localizing spot on a screen, which has at least the dimension of a grain in a photographic emulsion (10^{-7} m), and (2) the wavepacket is reduced in size whenever its “position” is observed. Nothing prevents us from taking over this argument, suitably adapted, when we consider the wavepacket as the electron itself. Notice also, that there is hardly any individual free elementary particle that is experimentally controlled for more than a fraction of a second. The assertion that it keeps together after it has left the apparatus goes far beyond any real experience. The wavepacket model may clash with some of our favored ideas on what an elementary particle is, but the decisive counter-argument could only be a clash with observed facts. We have not detected any such clash.

By the suitable adaptation of the argument mentioned above we mean the following. We recall that the wave function in a measurement turns into an eigenfunction (or a superposition of some eigenfunctions if the measurement precision is limited) of the operator that corresponds to the measurement apparatus. This is called the reduction or collapse of the wave function. It is not described by the Schrödinger equation, and we will discuss it in detail in Secs. 4.1, 4.5 and 4.6. In the case of a position measurement the (improper) eigenfunction of the position

operator in nonrelativistic quantum mechanics is a Dirac delta function. But due to some inaccuracy in the position measurement we may assume the function to have a finite though small extension. To such narrow functions the wavepacket is thus reduced in a position measurement. However, it is only in the Copenhagen interpretation that the reduction occurs in and only in a measurement. In our interpretation reduction is an objective physical process that occurs in certain physical situations, which need not be measurements. Thus we assume that the wavepacket when it traverses the successive screens suffers reduction to narrow functions not because its position is measured but because those objective reduction processes occur.

It is instructive to consider also non-free wavepackets which spread out. Consider an electron in an impenetrable macroscopic box. We may construct a wavepacket whose initial dimensions $\Delta x(0)$ and $\Delta p(0)$ are rather small, in particular $\Delta x(0)$ very small compared to the dimension of the box. This rather well localized packet may move back and forth between the walls, and it is a good model of a classical Newtonian particle. In the course of its oscillations, however, it will spread out and its width will become of the order of the dimension of the box, so that it ceases to be a good model of a classical particle; it is then a model of a classical wave swashing back and forth between the walls. Pictures of this behavior are shown by Brandt and Dahmen.⁽⁶⁰⁾ Another example is a small electron wavepacket orbiting around the nucleus in an atom, as in the planetary (Rutherford) model. Such a packet may be constructed by the superposition of some stationary eigenpackets, as suggested by Schrödinger.⁽⁵⁴⁾ However, as Heisenberg⁽³⁾ soon pointed out, the packet inevitably spreads out all around the nucleus in the course of time, although it remains within a finite radius.^(61,62)

In addition to the two objections, that ψ is complex and that free packets spread out, there are other, more serious ones. These include the objection that a wavepacket is said to be unable to account for the fact that “a whole particle” is always observed and never parts of it (dealt with in Sec. 2.3), that wave functions in multi-dimensional configuration space are used, and not only those in ordinary space (Sec. 3), and that the wave-function reduction, taken to occur instantaneously, allegedly cannot be a physical process (Sec. 4).

The subsequent sections are organized as follows. In Secs. 2 and 3 we develop the specific features of our interpretation, which go beyond the general principles of realism and of extended wavepackets discussed in Sec. 1, and which are different from those of the Copenhagen interpretation. In doing so we adopt an inductive approach selecting special physical situations to motivate and illustrate the postulates to be introduced. We conclude this with a summary of principles and postulates in Sec. 3.4.

Then, in Secs. 4 and 5 we turn to the application of these postulates to special problems. These sections show how the conceptual problems of the Copenhagen interpretation can be avoided, and by contrasting the two interpretations, they lead to a deeper understanding and a clarification of some additional points. Section 5.4 describes some crucial experiments.

Figure 1. The double-slit experiment

Finally, in Sec. 6 we show how the concepts of our interpretation applied to Schrödinger's configuration space with symmetrized functions lead us to the quantum statistical distribution and fluctuation formulas. The appendices contain the mathematical derivations of some important formulas in order to make the work self-contained.

2. SPECIFIC POSTULATES: SINGLE PARTICLES

2.1 The Double-Slit Experiment

We begin by considering the interpretation of the expression $|\psi(\mathbf{x}, t)|^2 d^3x$. Since the wavepacket is an extended object in our interpretation we can no longer speak of *the* position of the wavepacket, and the above expression can no longer be the probability of observing, at time t , the position within d^3x about \mathbf{x} , as in the Copenhagen interpretation. We shall introduce the new interpretation of $|\psi|^2 d^3x$ by considering the well-known double-slit experiment.

Consider a beam of electrons with average momentum p and little spread about this value. The beam is directed towards a wall, as shown in Fig. 1. The wall contains two parallel slits, which can be opened and closed. Behind the wall there is a detecting screen which registers the intensity $I(x)$ of the beam (number of electron counts per second) as a function of the distance x from the center O. If only slit 1 is open the intensity function $I(x)$ will look like I_1 ; if only slit 2 is open it will look like I_2 . If both slits are open the intensity function is not, however, the sum $I_1 + I_2$ but will look like I_{12} . The form of I_{12} is obtained simply by regarding the beam of electrons as a plane wave with a wavelength $\lambda = h/p$ and calculating the interference of waves originating from slit 1 and slit 2. These interference effects constitute the difference between I_{12} and $I_1 + I_2$.

The important point is that no matter how low the intensity of the incoming beam is made, the intensity function on the screen, when both slits are open, is always given by I_{12} provided we compensate for the lower incoming intensity by a longer exposure time to have the same total amount of energy (or total number of electrons) deposited on the screen. We may adjust the incident intensity until it is so low that it corresponds to one incoming electron per day. Thus, even a single

electron must correspond to a number of wave trains capable of interference with one another.

Let us first have a look at the Copenhagen interpretation of the double-slit experiment and the difficulties to which it leads. As is well known, in the Copenhagen school the expression $|\psi(\mathbf{x}, t)|^2 d^3x$ is not the probability that the particle at time t is in d^3x . Why can they not say this? Because they hold to the idea of a pointlike position and at the same time have to account for what is observed in the double-slit experiment. If the particle at every instant of time had a pointlike position this would mean that any single particle goes either through slit 1 or through slit 2 so that the intensity function on the screen would be $I_1 + I_2$ but never the observed I_{12} . Of course, one might invent a special type of motion of the particle, for instance some orbiting through both slits, before it finally arrives at the registering screen. It is, however, clear that this would have to be a quite complicated motion in order to explain the distribution I_{12} , and it would compete rather unfavorably with the simple explanation of interference of waves. Thus, if such constructions are left out of consideration the observed intensity distribution I_{12} leads to the conclusion that the particle does not have a sharp trajectory. In spite of this, the Copenhagen interpretation wants to continue, in some way or other, with the concept of a sharp position. The material point had already been a basic concept in classical mechanics^(63,64) and it had remained so in the first steps into the microworld, namely in Rutherford's and Bohr's models of the atom, which are planetary systems of point electrons.

Now, since the double-slit results forbid the particles to have a (sharp) position at all times, that position is ascribed to them only at the moment of measurement. Thus, "position" is no longer a permanent objective property of a particle but an "observable", which comes into existence only in the act of observation. If there is no observation a position must not be ascribed to the particle. This is not to be understood in the sense that the particle is extended when it is unobserved, and contracts to the point of observation when we observe it. In the Copenhagen interpretation the contraction happens only to the wave function, which is, however, not the particle itself (Sec. 4). As with position, other quantities like the components of momentum and angular momentum must not be ascribed to the particle, except at the moment of the respective measurement. All these quantities are "observables".

Now let us turn to the realistic interpretation. Here, the electron is a wavepacket, and its extension covers both slits, so that in any single passage both slits are involved and determine the final interference pattern. Only if the wavepacket has a transverse width (normal to the direction of its center) that is smaller than the distance between the slits will no double-slit interference effects be observed.

Let us have a closer look at how the intensity function $I(x)$ on the screen is obtained. The screen might be a photographic emulsion. The incoming electron beam causes black spots (after development) at certain positions on the screen. The intensity function $I(x)$ then is proportional to the spatial density of these spots. These spots are not points but have some extension. A spot is the result of a cascade

of processes which is initiated by the ionization of an atom of the emulsion by the incoming wavepacket. The localization is thus limited in practice by the extension of the black spot and in principle by the extension of the initiating atom or of the smallest emulsion wavepacket with which the incoming packet interacts. This does not mean that any distances that are smaller than the smallest wavepacket would have no meaning. We are only speaking about more or less *direct* position measurements. There are indirect measurements: a distance l may appear in a complicated equation, and by measuring all the other quantities in that equation the value of l may be calculated. In this way, for instance the formulas of quantum electrodynamics for leptons have been verified down to distances very much smaller than the diameter of any atom. The limitation in direct localizability does not present any basic difficulty in our interpretation, and we may assume in the present considerations that the black spots are small enough compared to the incoming wavepackets to be considered practically as points.

Then of course we can no longer say that a black spot at a certain position indicates that the electron is (or was) at this place. Rather, in our interpretation the packet, when traversing the special physical environment represented by the photographic emulsion, has induced a black spot at the observed position. The packet is like a cloud moving along while triggering thunder and lightning here and there. This leads us to the new interpretation of the expression $|\psi(\mathbf{x}, t)|^2 d^3x$.

2.2 The Action Probability

In our interpretation the expression $|\psi|^2 d^3x$ in some physical situations is also a probability, but different from the Copenhagen localization probability, and may be called an action probability. It is the probability that the wavepacket $\psi(\mathbf{x}, t)$ when placed into an appropriate physical environment, induces an observable effect within d^3x about \mathbf{x} during a small but otherwise arbitrary time interval Δt about t .

The interval Δt must only be short enough so that $|\psi|^2$ can be considered constant in it. The value of Δt does not appear because the expression $|\psi(\mathbf{x}, t)|^2 d^3x$ is meant to refer to an environment that is homogeneous, time independent, and such that the wavepacket with certainty induces an effect anywhere in it during Δt . This is what is meant by an appropriate physical environment. If the actual environment meets these requirements only partly the raw data must be corrected for this, and $|\psi|^2 d^3x$ refers to the corrected data. This is no new feature. Recall that Copenhagen quantum mechanics already presupposes perfect apparatus since $\int_{-\infty}^{+\infty} |\psi(\mathbf{x}, t)|^2 d^3x = 1$ there means that the object will be localized with certainty by the measurement at time t . By mentioning the time interval Δt we want to indicate that conceptually time plays the same role as space in our interpretation (Sec. 4.3).

Now, in the Copenhagen interpretation the expression $|\psi|^2 d^3x$ is defined as a probability, and accordingly the function ψ is defined as a probability amplitude. In our interpretation ψ is defined as a physical field, but the expression $|\psi|^2 d^3x$ is not *defined* as a probability. Rather we will show, by means of a calculation within the framework of the standard formalism, that $|\psi|^2 d^3x$ in certain physical

situations can be put numerically equal to our action probability.

The point might appear subtle, but it is important. Compare it with the following picture: Smoke comes out of a funnel. Let $\rho(\mathbf{x}, t)$ be the density of smoke, and let $p(\mathbf{x}, t)$ be the probability that a test body at \mathbf{x} and t will gather on its surface a certain amount of smoke particles in a certain time interval. This probability may be calculated on the basis of the physical laws involved in the process and may turn out to be proportional to the density ρ . Then, by referring ρ to some standard density and properly adjusting the chosen amount of gathered smoke particles, we can always arrange that the density $\rho(\mathbf{x}, t)$ is numerically equal to the probability $p(\mathbf{x}, t)$. Nevertheless, it is still smoke, and not probability, that comes out of the funnel. The expression $|\psi|^2$ in our interpretation plays the logical role of the smoke density ρ , not of the probability p , whereas in the Copenhagen interpretation $|\psi|^2$ plays the role of the probability p , and there is no expression in the formalism of quantum mechanics that in the Copenhagen interpretation would play the role of the smoke density ρ .

The calculation that will show that the expression $|\psi|^2 d^3x$ is numerically equal to our action probability will start with the well known transition probability formulas of standard quantum mechanics, such as⁽⁶⁵⁾

$$W_{a \rightarrow b} = |\langle b | U(t - t_0) | a \rangle|^2, \quad (1)$$

though with a different interpretation. In the Copenhagen interpretation $W_{a \rightarrow b}$ is the probability that if the system at the initial time t_0 is in the eigenstate $|a\rangle$, in a measurement at the later time t it will be found to be in the different eigenstate $|b\rangle$.⁽⁶⁶⁾ We may consider formula (1) in the Schrödinger representation and particularly in one where the position variables are diagonal⁽⁶⁷⁾

$$W_{a \rightarrow b} = \left| \int_{-\infty}^{+\infty} \psi_b^*(\mathbf{x}, t) U(t, t_0) \psi_a(\mathbf{x}, t_0) d^3x \right|^2.$$

The state vectors $|a\rangle$ and $|b\rangle$ are here replaced by the wavepackets ψ_a and ψ_b , and $W_{a \rightarrow b}$ can now be considered as the probability that the wavepacket changes its form from the form ψ_a to the form ψ_b . The point is that we now can speak of probabilities without already committing ourselves to the probability interpretation of ψ . It is irrelevant whether ψ_a and ψ_b are themselves probability amplitudes or physical fields; the transition probability refers only to the occurrence of the transition, not to the thing that makes the transition. Also, in the Copenhagen interpretation there is on the one hand the system (particle) and on the other hand the states (wavepackets), and one might imagine the particle jumping from one state into the other, like a traveller who changes trains. Nothing of this, of course, can be said in the realistic interpretation; here all that exists is the wavepacket changing its form.

In the realistic interpretation it is most important that the transition need not be connected with any measurement or observation; here we have no postulate automatically guaranteeing that the transition reveals itself to any observer. Instead,

for any particular quantity to be measured, we have to show that transitions exist that are accompanied by processes that lead to detectable effects, and that the probability of these effects is given by our action-probability formula. Since we do not insist that $|\psi|^2 d^3x$ be equal to an action probability under all circumstances, it suffices to select transitions that refer to those physical processes that are actually employed when the expression $|\psi|^2 d^3x$ is subjected to an experimental test. And it is sufficient when these transitions accomplish their task only approximately, or under special conditions. This procedure clearly is close to what physicists really do when they are going to measure a particular quantity.

In principle, the effect induced by the wavepacket ψ need not be observed. In realism, if an effect occurs, it occurs independently of our observing it. And it may be any effect. It is only because we want to refer to the same physical situation as the Copenhagen interpretation that it must be an observable effect, observable according to the criteria of the Copenhagen interpretation. This means merely another physical condition on the considered transitions in our interpretation, i.e. on the admitted final states ($|b\rangle, \psi_f, \Phi_n$, below) in these transitions, or on the “appropriate” physical environment. A stable macroscopic fixation, memory or record of the effect will be required, but how to achieve this is not a specific problem of quantum mechanics and does not concern us here.

As a specific physical process for the interpretation of $|\psi|^2 d^3x$ we choose the formation of a black spot in a photographic emulsion induced by an electron. We use the fact that the probability of producing a black spot in d^3x is proportional to the probability of ionizing an atom in d^3x , which in turn is proportional to the probability that the electron wavepacket makes a transition from its initial form ψ_i to some final form ψ_f , simultaneously with an atom in d^3x making a transition from the ground state Φ_0 to some ionized state Φ_n . The calculation can proceed entirely within the framework of the standard formalism. We only reverse the logical order: in the Copenhagen interpretation the localization probability is a primary and the transition probability a derived quantity, whereas in our interpretation it is the other way round.

Moreover, since the chosen processes are all well known, and are consistently dealt with in standard quantum mechanics, the production of a black spot in d^3x calculated (1) via the transition probabilities and (2) by means of the expression $|\psi(\mathbf{x}, t)|^2 d^3x$ must of course lead to the same result. It is, nevertheless, instructive to see how this works in detail. Actually, the derivation will only be possible under the special physical conditions that the incoming electron wavepacket is large compared to an atom and that the range of the interaction between the wavepacket and the atom is effectively short. These conditions are, however, always satisfied when the localization-probability formula $|\psi|^2 d^3x$ is submitted to a direct experimental test. In order not to interrupt the line of argument we do not carry out the calculation here but postpone it to Secs. 4.2 and 4.3 where we discuss measurement in general.

Although the black spot on the screen in our interpretation does not mean that the particle “had been there”, it still means that it had been somewhere around,

in the sense that it had covered the position of the spot. For example, suppose we know the shape of the wavepacket but not its absolute location. We can then calculate, by maximum-likelihood methods, the probability that the region of the black spots includes the center of the packet. Or suppose the other case, that we know the position of the center of the packet but not its shape. We may then measure the shape in the following manner: let the direction of the wavepacket's center be the y direction, put a screen in the x - z plane in its way, and register the black spot appearing on the screen. Repeat this under identical initial conditions. The distribution of spots on the screen then images the shape of the wavepacket, or, more precisely, its projection on the x - z plane. And the width Δx of the distribution in x direction is the width of the wavepacket.

After the wavepacket has induced a black spot it will be changed to some extent, simply on the basis of the principle that action equals reaction. This change implies that the production of a black spot in the above-mentioned measurement of the packet's shape refers to the production of the *first* spot in each of the respective repetitions. The incoming wavepacket may induce more than one black spot, particularly if the screen is thick; but after it has induced the first one its form is changed in a way that is usually too complicated to calculate, and there is no longer a well-defined initial condition for the second black spot, etc. Therefore we need the repetition of the experiment under identical initial conditions; in other words we need an ensemble of equal wavepackets.

The interpretation of $|\psi(\mathbf{x}, t)|^2 d^3x$ as an action probability in ordinary position space can be extended in an analogous way to other spaces. Let $\varphi(a, t)$ be the wave function in some "A space", in the sense of general transformation or representation theory.⁽⁶⁸⁾ The function $\varphi(a, t)$ is interpreted as the wavepacket in A space, and it may be an extended object in A space as it is in position space. Since we say that the wavepacket with probability $|\psi(\mathbf{x}, t)|^2 d^3x$ induces an effect in d^3x about \mathbf{x} , or acts at \mathbf{x} in position space, we may say that the wavepacket with probability $|\varphi(a, t)|^2 da$ acts in da about a in A space. This also applies to discrete eigenvalues a_n in which case $|\varphi(a_n, t)|^2$ is the probability of acting at the point a_n in A space. Again, for this interpretation to be possible the wavepacket must be in an appropriate physical environment, and the action occurs during a short but otherwise arbitrary time interval.

As an illustration consider a massive-particle wavepacket $\tilde{\psi}(p, t)$ in one-dimensional momentum space. In order to determine $|\tilde{\psi}(p, t)|^2$ experimentally we let the wavepacket collide head on with a photon (Compton effect⁽⁶⁹⁾). By preparing the initial and measuring the final momentum (frequency) of the photon the initial momentum p of the particle can be calculated via the conservation laws of energy and momentum. In any repetition of this collision under identical initial conditions the result will be a different value of p ; in the different collisions the packet acts with different values in momentum space. In order to determine the shape $|\tilde{\psi}(p, t)|^2$ of the wavepacket in momentum space we must use an ensemble, say a beam, of equal packets and observe many collisions. This corresponds exactly to the above-described procedure of determining the shape of the packet in ordinary space. And

the above condition that the atom (black spot) must be small compared to the size of the wavepacket now corresponds to the condition that the width of the photon packet in momentum space $\Delta\nu h/c$ (both before and after the collision) must be small compared to the width Δp of the massive-particle wavepacket.

2.3 The Internal Structurelessness of Wavepackets

We will consider now more closely how the black spot in an emulsion screen is brought about. This will lead us to perhaps the most remarkable property that we have to ascribe to a wavepacket. Consider a beam of electromagnetic radiation falling onto a screen. Let us first treat the beam classically as a continuous field with the time-averaged energy flux density $\bar{S} = \epsilon_0 c \overline{E^2}$ in W/m², say. The registering screen, on the other hand, is conceived to consist of atoms in the sense of quantum mechanics. The incoming radiation will then cause black spots at certain places on the screen, and their density classically is proportional to the energy flux density of the radiation at these places. The resulting pattern will exhibit a granular structure but this does not demonstrate that the incoming radiation has particle or quantum properties; it only shows that the screen has, which we have presupposed anyway. The wind is not made of particles only because it causes an integral number of trees to fall in the forest, as Marshall and Santos put it.⁽⁷⁰⁾ To see what may be called the quantum or particle aspect of the incoming radiation we recall that the black spot is the result of a cascade of processes which is initiated by the ionization of an atom. This atom, in order to become ionized needs some threshold energy E_{thr} . Of course, knowing quantum mechanics we assume that $h\nu \geq E_{\text{thr}}$, but this is not sufficient for ionization in the classical picture. Imagine that the atom gets this energy by absorbing and accumulating all the energy of the incoming classical radiation which arrives on its area σ , like a dust collector that is put in a stream of polluted air. With the energy flux density of the radiation \bar{S} the energy accumulated by the atom during the time t is $E = \bar{S}\sigma t$. When the atom has accumulated the energy E_{thr} it becomes ionized and initiates the formation of the black spot. The accumulation time needed for this is

$$t_{\text{acc}} = E_{\text{thr}}/(\bar{S}\sigma). \quad (2)$$

Now, the value for t_{acc} according to the above picture turns out to be of the order of hours or years in situations where effects are actually observed immediately after the arrival of the radiation. For example, in the interference experiments of Reynolds et al.⁽⁷¹⁾ the average light intensity in the interference pattern produced by a Fabry-Perot interferometer corresponded to as little as 1 photon/(mm²s). They used light of wavelength 5.358×10^{-7} m which means photons of energy 2.18 eV. In the classical picture this corresponds to $\bar{S} = 3.5 \times 10^{-13}$ W/m². Formula (2) with an atomic area of $\sigma = 10^{-18}$ m² and an assumed threshold energy of $E_{\text{thr}} = 2.18$ eV then gives an accumulation time of $t_{\text{acc}} = 10^{12}$ s = 30000 years! The interference pattern would thus only appear after years, but then it would appear fully in one flash. Actually the interference pattern was obtained in two minutes, and in various runs with reduced exposure time the authors verified that the pattern is built up gradually as time proceeds. The first black spots are induced

immediately after the arrival of the radiation (cf. also Sec. 2 of Ref. 72). Of course the total energy absorbed by all atoms during the whole accumulation time is the same in both cases. The difference is that classical theory would have it absorbed all in the last moment (at t_{acc}), whereas experiment shows that it is absorbed in many small portions distributed over the accumulation time.

These facts suggest that the radiation cannot be treated as a classical continuous field. Even with somewhat different assumptions one arrives at the same conclusion. For example, if we assume the area σ to be as large as 10^{-14} m^2 we still obtain an accumulation time of three years. Also, other mechanisms of energy collection considered by Lorentz and by Sommerfeld and Debye gave shorter but still far too long accumulation times.⁽⁷³⁾ An alternative hypothesis, first put forward by Lenard in 1902, was that the energy E_{thr} is already stored in the atom and the radiation only triggers the release of an electron. This hypothesis turned out to be contradicted by several experimental observations, the most serious of these being the discovery that the electron emission is always accompanied by strong absorption of the incoming radiation. The hypothesis was thus abandoned by scientists and by Lenard himself in 1911.⁽⁷³⁾

It is thus safe to conclude that the energy for ionizing the atom is not the energy contained in the cylinder that the atom has cut out of the field up to the moment of its ionization. Rather, the energy must come from a much larger volume. We might now assume that the total energy classically contained in this larger volume actually is already concentrated in a point somewhere within that volume and that this point happened to hit the atom. However, this would lead us to the difficulties in explaining the double-slit interferences, and therefore we reject this assumption. Rather, the energy that ionizes the atom must be transferred from the outer regions of that volume to the point where the atom is. And since the classical field intensity may be arbitrarily low the volume may be arbitrarily large and the energy transfer has to occur with superluminal velocity. Taking this conclusion seriously, I want to integrate it into the theory in the following manner: The field consists of extended wavepackets, each of sufficient total energy, and it is a wavepacket that ionizes the atom. In this process the wavepacket acts as a whole. The transfer of energy occurs with superluminal velocity but is restricted to the interior of a wavepacket. And this I conceive to hold not only for electromagnetic wavepackets, i.e. photons, but for all types of packets, i.e. electrons, protons etc. Any matter and radiation field thus has a granular structure.

This is not the first time that such a proposal is made. In his paper on the classical theory of radiating electrons Dirac supposes the electron (though not its charge) to have a finite size, of the order of the classical electron radius, and he writes:

it is possible for a signal to be transmitted faster than light through the interior of an electron. The finite size of the electron now reappears in a new sense, the interior of the electron being a region of failure, not of the field equations of electromagnetic theory, but of some of the elementary properties of space-time.⁽⁷⁴⁾

Superluminal propagation of information inside extended particles has also been considered by Vigier and Cufaro-Petroni.^(75,76)

In fact I would consider the transmission of effects inside the wavepacket as occurring not only with superluminal but with infinite velocity. However, I would propose a different viewpoint: the wavepacket is a really elementary object. It represents a fundamental region of space that must in no way be subdivided any further; if we were to do that it would be the smaller parts that would play the role of fundamental regions. Therefore it is not that the velocity is infinite, rather that there is no distance to travel. All points in the interior have to be identified, in a sense they are all one and the same point. The interior of any wavepacket is structureless, in the sense that there are no parts between which energy or signals of any kind may be exchanged. I call this the “internal structurelessness” of wavepackets. Other terms may be internal spacelessness, pointlikeness, rigidity, or telepathy.

What was said about energy applies equally to charge, magnetic moment, rest mass etc. All qualities are ascribed to the wavepacket as a whole. This accounts for the point-like behavior of the wavepacket in the interaction. The Copenhagen school interprets this by assuming that besides the wave function there is a particle, a point particle, and this point carries all the quantities with it. This avoids the internal structurelessness but only at the price of the wave-particle-duality problem.

The wavepacket, in order to maintain internal structurelessness, needs not be spatially connected but may consist of nonoverlapping parts. This is evidenced, for example, in the situations of the two slits, the Stern-Gerlach magnet and the single-crystal neutron interferometer.⁽⁷⁷⁾

Quantities like momentum and energy (of a free wavepacket) have no single fixed value, and the wavepacket may act with different values of them in different interactions. Nevertheless, in no case does this mean that only spatial parts of the packet were involved in the interaction.

All this refers to wavepackets that represent single elementary particles. In Secs. 3.1 and 3.2 we shall consider wavepackets that represent condensed particles, and these may act with rational fractions of some of their quantities, although not of their space.

The absence of structure comes into play only when the pointlike features of the interaction as such are considered, not when the probability of its occurrence is calculated. In this calculation, via the action probability or via the transition probability, the full $\psi(\mathbf{x}, t)$ is used, defined for every point in the wavepacket. And the probability of occurrence of the effect not only refers to occurrence in x space but also in momentum space and other spaces, as well as to correlations between effects in various spaces. Thus, for example, when a photon wavepacket goes through a prism, behind the prism the packet spreads out considerably, and at some places it acts preferably with high momenta (blue color) and at others with low momenta (red color). Or when an (atomic) electron wavepacket passes through a Stern-Gerlach-type apparatus, at some places it acts as a spin-down

electron and at others as a spin-up electron. Nevertheless, these effects are still to be counted among the external effects and have nothing to do with internal structurelessness. The internal structurelessness in these cases manifests itself in the fact that the photon wavepacket totally disappears after having shown up at a certain place, and that the electron packet shows up with all its spin at the place where it acts and is then reduced to a pure spin-up or a pure spin-down packet, respectively. Also, the internal structurelessness has nothing to do with the structure of the proton (form factor) as revealed in the scattering experiments of Hofstadter,⁽⁷⁸⁾ which is to be compared to the structure of the ψ function of the atomic electron in the Stern-Gerlach experiment.

The wavepacket in a sense appears structured when seen from outside but unstructured when seen from inside. This is why I speak of *internal* structurelessness and not just of structurelessness. I see no other way to cope with the observed reality than to first introduce $\psi(\mathbf{x}, t)$ for all values of \mathbf{x} and t and to then restrict the meaning of \mathbf{x} and t inside the wavepackets through additional postulates when internal relations of any kind are involved. In any case some postulate of this kind is inevitable if one wants to deal with fundamental objects that are both extended and indivisible.

Think of a thunder cloud that moves over a flat land covered with trees. Lightning will strike the Earth at that point (x, y) for which the threshold conditions of the discharge of the cloud are satisfied for the first time. These conditions, in a simplified model, are determined by the electric potential of the cloud “ $|\psi(\mathbf{x}, t)|^2$ ” and the height of the trees $\rho(x, y)$, but if we suppose that all trees are of almost equal height ($\rho(x, y) \approx \text{const.}$), with only small fluctuations due to the wind, the probability of lightning is essentially determined by $|\psi(\mathbf{x}, t)|^2$. On the other hand, once the lightning occurs at some point (x, y) , the whole cloud will be discharged there.

The internal structurelessness implies that the wavepacket in a sense behaves like a rigid body of classical (Newton-Euler) mechanics: imagine a large cigar-shaped wavepacket that interacts with another, small wavepacket at one end. The large (as well as the small) wavepacket is modified by this interaction (imagine that it was white and turns grey), but the modification must not be imagined as starting from the one end and gradually expanding over the whole packet, rather the packet is simultaneously modified as a whole (in some Lorentz system, the rest system or the center-of-momentum system, say. In another system the simultaneous events would move with the (superluminal) phase velocity of de Broglie waves). The modification as such need not occur instantaneously but may take some time.

In a second interaction with another small packet at the other end the large packet is already modified (at least to some extent) and the characteristics of the second interaction are thus influenced by the first. The time between the two interactions may be shorter than the time for a light signal to travel from one end to the other, so that the two correlated interactions occur with a spacelike separation in space-time. This means some kind of nonlocality.

The interactions may or may not be accompanied by reductions of the wave-

packet(s). It may happen that the first interaction/reduction also leads to a contraction of the cigar-shaped wavepacket to a smaller spatial region. This would make the above situation a bit more complicated but it would still allow us to conceive two interactions with spacelike separation. Of course the wavepacket must not contract to a point, but this is generally excluded in our interpretation. The infinite velocity or spacelike separation of correlated events in this situation does not lead to a contradiction with relativity theory, as will be shown in Sec. 5 on the Einstein-Podolsky-Rosen problem.

An important corollary of the internal structurelessness is the absence of any self-interaction. Since there are no parts of a wavepacket there can be no interactions between them. This does not affect the construction of a wavepacket by the superposition and interference of plane waves or other orthogonal systems of functions, which is a mathematical treatment; but it is very relevant to, for example, the Coulomb interaction between parts of a charged particle, or the action of radiation emerging from one part on another part of the same particle. The first kind of interaction leads to the infinite self-energy of a point particle. The postulate of internal structurelessness excludes this. The absence of self-interaction of an electron is an old question and has often been conjectured, especially in connection with the diverging self-energy and the stability of the classical electron.^(79,80) As early as 1909 Lorentz, the theoretical father of the electron, stated:

In speculating on the structure of these minute particles we must not forget that there may be many possibilities not dreamt of at present; it may very well be that other internal forces serve to ensure the stability of the system, and perhaps, after all, we are wholly on the wrong track when we apply to the parts of an electron our ordinary notion of force.⁽⁸¹⁾

In fact, we may even regard the success of the Schrödinger equation as confirmation of the absence of self-interactions.⁽⁸²⁾ Consider the Schrödinger equation for an electron in an electrical potential $V(\mathbf{x}, t)$

$$i\hbar \frac{\partial \psi(\mathbf{x}, t)}{\partial t} = -\frac{\hbar^2}{2m} \Delta \psi(\mathbf{x}, t) - eV(\mathbf{x}, t) \psi(\mathbf{x}, t). \quad (3)$$

Let us for the moment regard the quantity $\rho(\mathbf{x}, t) := -e|\psi(\mathbf{x}, t)|^2$ as the charge density of the electron, and let us write the potential $V(\mathbf{x}, t)$ as the sum of two terms

$$V(\mathbf{x}, t) = V_o(\mathbf{x}) + V_e(\mathbf{x}, t), \quad (4)$$

where $V_o(\mathbf{x})$ is the potential produced by the atomic nucleus (proton) plus some outside charges, and $V_e(\mathbf{x}, t)$ is the potential produced by the charge distribution of the electron itself. This can be written as

$$V_e(\mathbf{x}, t) = \frac{1}{4\pi\epsilon_o} \int \frac{\rho(\mathbf{x}', t)}{|\mathbf{x} - \mathbf{x}'|} d^3x'. \quad (5)$$

$V_e(\mathbf{x}, t)$ represents some action of the electron on itself. Inserting (4) and (5) into (3) leads us to the nonlinear integro-differential equation

$$i\hbar \frac{\partial}{\partial t} \psi(\mathbf{x}, t) = -\frac{\hbar^2}{2m} \Delta \psi(\mathbf{x}, t) - eV_o(\mathbf{x})\psi(\mathbf{x}, t) + \psi(\mathbf{x}, t) \frac{e^2}{4\pi\epsilon_o} \int \frac{|\psi(\mathbf{x}', t)|^2}{|\mathbf{x} - \mathbf{x}'|} d^3x', \quad (6)$$

which differs from the familiar Schrödinger equation by the last, self-interaction term. On the other hand, it is the familiar Schrödinger equation, and not Eq. (6), that gives the correct results, for example for the hydrogen eigenfunctions.

We note, however, that the absence of interactions between spatial parts of a wavepacket does not exclude the existence of recoil effects of the emitted radiation on the emitting wavepacket as a whole, as it is considered by Barut and collaborators in their approach to quantum electrodynamics without canonical quantization.⁽⁸³⁾

2.4 Normalization and Quantization

It follows from the Schrödinger equation with a Hermitean Hamilton operator that the quantity

$$I = \int_{-\infty}^{+\infty} \rho(\mathbf{x}, t) d^3x$$

with

$$\rho = |\psi(\mathbf{x}, t)|^2 \quad (7)$$

is independent of time. In the Copenhagen interpretation the integral is the probability of obtaining any value for the position of the particle whose wave function is $\psi(\mathbf{x}, t)$ in an appropriate position measurement. The probability assumes a 100% efficiency of the apparatus (Sec. 4.3), hence the value of the integral has to be 1. Thus, the conservation of I is in accordance with the fact that in nonrelativistic quantum mechanics a particle can neither be created nor destroyed (“conservation of probability”).

In the realistic interpretation the question is what the conserved quantity $I = \int \rho d^3x$ is, if not a probability, and why it should be equal to 1. Let us first consider the question what the field $\psi(\mathbf{x}, t)$ is. This is easy: the field is the field. It is an irreducible entity, a “primitive notion”, in the same sense that the electromagnetic field is. Originally the electromagnetic field had been “explained” by the ether, but for well known reasons the ether has been dismissed and the field has been established as a quantity in its own right (Einstein⁽⁶³⁾). We become familiar with the field by elaborating how it behaves, that is, by finding rules that connect it with well-known phenomena. ψ is just a “matter” field (electron field, proton field etc.) as \mathbf{E} is the “electric” field. As Jammer puts it⁽⁸⁴⁾: “physical reality is what it is because it does what it does.”

Likewise, the meaning of the integral $I = \int \rho d^3x$ depends on what role I , or ρ , plays in the formulas describing familiar phenomena. In quantum field theory the

integral I , with some more general ρ than just (7), is identified with the operator for particle number, or for particle number minus antiparticle number, or for total charge, either electric or baryonic or leptonic etc. I would like to leave the meaning of I open. This does not pose a problem for a realistic interpretation. What matters here is that I a priori is not a probability.

In the Copenhagen probability interpretation it is important that ρ is a (positive) definite quantity. There are, however, some otherwise perfectly reasonable equations that do not lead to a positive definite ρ . Let us consider this point more closely. From a more general point of view the quantity $\rho(\mathbf{x}, t)$ is a density in a continuity equation

$$\partial\rho/\partial t + \text{div}\mathbf{j} = 0. \quad (8)$$

In the relativistic theory ρ and \mathbf{j} in (8) form a four vector, the continuity equation expresses the vanishing of its four divergence, and $I = \int \rho d^3x$ is a Lorentz scalar. The Klein-Gordon equation thus leads to

$$\rho = \frac{i\hbar}{2m} \left(\psi^* \frac{\partial\psi}{\partial t} - \frac{\partial\psi^*}{\partial t} \psi \right), \quad (9)$$

which is not positive definite. This was the main reason why Dirac developed his equation. The Dirac equation leads to

$$\rho = \sum_{s=1}^4 |\psi_s|^2, \quad (10)$$

which is positive definite. As we ascribe the same status of reality to the matter and to the radiation field we should also include the photons in this consideration. Here a conserved Lorentz scalar for the source-free electromagnetic field is, for example,^(85,86)

$$\hat{I} = \frac{1}{8\pi c} \int d^3x \phi^\dagger \frac{1}{p} \phi \propto \int d^3x d^3x' \frac{\phi^\dagger(\mathbf{x}) \phi(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^2} \quad (11)$$

with $\phi = (E_x - iB_x, E_y - iB_y, E_z - iB_z)$. The density

$$\rho = \frac{1}{8\pi c} \phi^\dagger \frac{1}{p} \phi \quad (12)$$

is not local as seen from the second integral in (11). Therefore this density has been dismissed as “unnatural” by Pauli.⁽⁸⁷⁾ The density (12) is also not positive definite.

In the realistic interpretation ρ is no longer defined as a probability density. So, here is it completely irrelevant whether or not ρ is positive definite. Also, a nonlocal ρ , such as (12), is quite acceptable in this theory where the internal structurelessness is already a patently nonlocal property. The integral $I = \int \rho d^3x$

can always be considered as extending over the volume of at least one wavepacket and thus can always represent a good physical quantity. All that we need are formulas for cross sections that are in accordance with observation. Indeed, in some places in quantum field theory the postulate of a positive definite ρ had already been abandoned.^(88–92)

Setting $I=1$ for any ψ then means that the conserved quantity I (whatever it is) has the same value for any wavepacket. The value 1 as such is not essential, it may be any other number; the essential point is that it is always the same number.

Setting $I=1$ may be interpreted by saying that a wavepacket represents one quantum or one field quantum; even without in any way specifying the nature of that quantum and without excluding the possibility that a particular kind of wavepackets (photon packets, electron packets etc.) represents a particular kind of quanta. This would be in accordance with the view that the wavepacket always acts as a whole, as explained in the preceding section. Thus we postulate that *any wavepacket represents one quantum*. Actually, this postulate will be extended in Sec. 3, in that a wavepacket may represent any integer number of quanta, although it will remain normalized to unity, according to the standard formalism.

This postulate in realism is a true “quantum” postulate, going beyond classical continuum physics. The Schrödinger equation for $\psi(\mathbf{x}, t)$ as it stands is still a classical field equation.⁽⁹³⁾ The appearance of the typical quantum constant \hbar in it only indicates the kind of fields to which the equation refers, namely to Schrödinger, de Broglie or matter fields whose characteristic wave quantities ω and λ are related to the particle quantities E and p by the relations $E = \hbar\omega$ and $p = \hbar/\lambda$. Even the special boundary conditions introduced in order to get “quantized” solutions, which can be distinguished from one another by parameters that take on discrete (eigen)values, are not enough to provide a genuine quantum character. The energy eigenvalues in the hydrogen atom, for example, follow already from the normalizability postulate $I < \infty$. The postulate $I=1$ goes beyond this. Eigenvalues already appear in classical macroscopic physics, and it is significant that the title of Schrödinger’s famous papers was “quantization as a problem of eigenvalues”.

2.5 The Heisenberg Relations

The Heisenberg relations

$$\Delta x \Delta p_x \geq \hbar/2 \quad (13)$$

play a central role in the continuing discussions on the physical meaning of quantum theory. Many different interpretations have been advanced,⁽²⁴⁾ and it is not our intention to review them here. Rather, we shall pick out a few points that serve to clarify our interpretation.

In the realistic interpretation the ranges Δx and Δp_x are simply the extensions of the wave packet. Δx is the extension (in x direction) in ordinary space, $\Delta x = \langle (x - \langle x \rangle)^2 \rangle^{1/2}$, for example, and Δp_x is the extension in momentum space. A direct measurement of Δx may be effected by measuring the scatter in the spatial distribution of the black spots induced by an ensemble of equal wavepackets on a screen, and Δp_x is obtained in an analogous manner, as described in Sec. 2.2.

Both Δx and Δp_x are simultaneously ascribed to a wavepacket. The Heisenberg relations (13) thus express a relation between these ranges.

The derivations of the Heisenberg relations presented in most textbooks consider operators for position and momentum and their Hermiticity and noncommutativity, and the Schwarz inequality for the state vectors. This might suggest that the Heisenberg relations are basic structural features of the formalism of quantum mechanics. We do not think so. The relations can also be derived by speaking only of wavepackets.⁽⁹⁴⁾ Remember the relation $\Delta x \Delta \nu \geq c/4\pi$ and its interpretation for a pulse of classical electromagnetic radiation, known long before the advent of Heisenberg's relation.

The importance of the Heisenberg relations stems from the importance of the wavepacket nature of the elementary particles. In fact, one may characterize quantum mechanics as the theory of those phenomena where the concept of a point particle or point localization irreparably breaks down. In the regime of classical mechanics only the first moment of the function $|\psi(\mathbf{x}, t)|^2$, namely, the center value $\langle \mathbf{x} \rangle$ plays a role (Ehrenfest theorem), whereas in the proper quantum regime the higher moments come into play. Already the first step, the inclusion of only the second moments Δx into consideration, marks the essential differences from classical mechanics.

Only in those physical situations where all relevant lengths are large compared with the dimensions of the wavepacket will the quantum theory lead to the "classical" results. This is essentially what is behind the usual statement that the "classical limit" means $h \rightarrow 0$ or $\lambda \rightarrow 0$ where λ is the de Broglie wavelength. Consider a one-dimensional wavepacket with a momentum width Δp such that $\Delta p/\langle p \rangle = \text{const} \ll 1$. A beam made up of such wavepackets would be quasi-monochromatic, and any beam to which an energy is at all ascribable ("a 45 keV electron beam") must be of this type. The Heisenberg relation for Gaussian minimum packets then reads $\Delta x = 1/(4\pi \text{const}) \times h/\langle p \rangle$ from which it is seen that, other things remaining constant, $h \rightarrow 0$ or $\lambda = h/\langle p \rangle \rightarrow 0$ is equivalent with $\Delta x \rightarrow 0$. For other than Gaussian packets only different factors appear in the relation between Δx and $h/\langle p \rangle$. However, Δx may still remain large even when h or λ goes to zero because even for Gaussian packets the Heisenberg relation holds only at one instant of time. Thus the limits $h \rightarrow 0$ or $\lambda \rightarrow 0$ need not always lead to classical behavior; only $\Delta x \rightarrow 0$ does so.

Since the Copenhagen interpretation insists that a sharp position, or a sharp momentum of the particle can be obtained (at least in principle) as a result of a measurement, the Heisenberg relations are called uncertainty, inaccuracy or indeterminacy relations. In the realistic interpretation here presented, there is no sharp position to be uncertain about. Asking for any sharp value of position or momentum of a particle is asking the wrong question. Once we are given the mathematical form of the wavepacket (which includes its ranges) no uncertainty is left concerning this packet. Therefore we do not speak of uncertainty relations; rather we speak of complementarity or reciprocity relations, or just Heisenberg relations. One may say that there is still the uncertainty about the place where

the wavepacket will induce an effect. This uncertainty depends, however, on the environment and is not a characteristic of the packet alone.

While in a realistic interpretation the Heisenberg relations a priori have nothing to do with any measurement (cf. Planck⁽⁹⁵⁾ and Bunge and Kalnay^(28,26)), in the Copenhagen interpretation it is to the contrary. Thus, Bohm writes:

If a measurement of position is made with accuracy Δx , and if a measurement of momentum is made *simultaneously* with accuracy Δp , then the product of the two errors can never be smaller than a number of order \hbar .^(96,97)

This suggests that the widths Δx and Δp_x were primarily those of the measurement apparatus and not those of the wavepacket, which is also suggested in the example given by Messiah⁽⁹⁸⁾ where the Δ 's are widths of slits in diaphragms. Also, in the framework of the Copenhagen interpretation efforts have been made to find a physical mechanism by which the Heisenberg relations are brought about, and consideration has been given to the influence of the measuring devices on the measured object. Bohr describes physical situations where the experimental arrangement suited for measuring the exact positions excludes the experimental arrangement suited for measuring the exact momenta of some particles, and he writes:

As repeatedly stressed, the principal point is here that such measurements demand mutually exclusive experimental arrangements.⁽⁹⁹⁾

I do not think that this is a principal point, characteristic of situations in quantum mechanics. In classical physics such situations exist as well. Consider a sugar cube and the two properties of being soluble and being burnable. Evidently, measurements of these properties demand mutually exclusive experimental arrangements. Yet nobody would find anything unusual in this situation, and nobody would have any objection to saying that the cube is burnable and soluble. In the normal realistic language simultaneous existence has nothing to do with simultaneous measurability. The justification for calling the individual cube burnable and soluble rests on the fact that one has at one's disposal many cubes that are presupposed to be equal with respect to the considered properties. It is easy to imagine cubes made up of a mixture of several substances so that, depending on the mixing ratio, the cubes are easier to burn but more difficult to dissolve, and the degrees of solubility and combustibility are in a certain reciprocal or complementary relation to each other, in analogy to Δx and Δp_x . Similar examples have been described by Janossy⁽²³⁾ and by Koopman.⁽¹⁰⁰⁾ Indeed, any procedure of destructive testing of materials can provide new examples.

The relations (13) may be generalized to any two quantities which are represented by Hermitean operators A and B ^(56,101)

$$\Delta A \Delta B \geq \frac{1}{2} |\langle [A, B] \rangle|, \quad (14)$$

and the realistic interpretation is again that these relations and the noncommutativity of the operators express correlations between properties (the ranges ΔA and

ΔB) of wavepackets. This interpretation is independent of whether the operators have discrete or continuous eigenvalues. For example, the relation between the components of angular momentum

$$\Delta l_x \Delta l_y \geq \frac{1}{2} \hbar |\langle l_z \rangle|, \quad (15)$$

means that the particle which is an eigenpacket of $l_z (\neq 0)$ cannot at the same time be an eigenpacket of l_x and l_y , rather it has the finite ranges Δl_x and Δl_y . When the same packet in an appropriate physical situation changes into an eigenpacket of l_x , say, then its range Δx shrinks to zero while it acquires the finite ranges Δl_y and Δl_z .

3. SPECIFIC POSTULATES: SYSTEMS OF PARTICLES

3.1 Similar Wavepackets: Coalescence and Splitting

In the preceding section the concepts of a realistic interpretation concerning single particles were set out. In the present section we shall extend our considerations to include systems of particles.

Many-particle systems in quantum mechanics are described by configuration space wave functions $\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N, t)$. This does not prevent us from maintaining a realistic interpretation, in analogy to Sec. 2.2 above: the variable \mathbf{x}_1 refers to wavepacket number 1, \mathbf{x}_2 to wavepacket number 2 and so on. The wavepackets have their properties even when not observed, and the mathematical expression

$$|\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N, t)|^2 d^3 x_1 d^3 x_2 \dots d^3 x_N,$$

for example, which in the Copenhagen interpretation is taken to be the probability that particle 1 is observed at \mathbf{x}_1 , particle 2 at \mathbf{x}_2 , and so on, in the realistic interpretation is the probability that wavepacket 1 induces an effect at \mathbf{x}_1 , wavepacket 2 at \mathbf{x}_2 , and so on. This holds generally, independent of the more specific features of the interpretation.

It is of particular importance to distinguish between similar and dissimilar wavepackets. These two classes are those that usually are called identical and non-identical particles (although not in Dirac's book). One reason for our not following the common terminology is the ambiguity in the meaning of the word identical. We may say "Lord Kelvin and William Thompson are identical", which means that the two are one and the same person; but we may also speak of identical twins, which means two different persons. Another reason for our terminology will be mentioned below. The distinction between similar and dissimilar wavepackets is important because in our interpretation similar wavepackets can *coalesce* but dissimilar wavepackets cannot. We begin by considering similar wavepackets. Dissimilar packets will be considered in Sec. 3.3.

In quantum mechanics the effects of similarity of particles go beyond those met in classical mechanics in both a mathematical and a conceptual way.

On the mathematical side the Hamilton operator H (in fact any operator and probability expression) of a system of similar particles is invariant under all permutations of the particle labels (coordinate subscripts), including the spin-component labels: $[P, H] = 0$, with $P =$ any permutation operator. This is nothing more than the permutation invariance of the Hamilton function of classical mechanics translated into quantum-mechanical language. It implies that if Ψ is a solution of the Schrödinger equation $P\Psi$ is also a solution, and that Ψ can be written as a linear combination of functions transformed by the irreducible representations of the permutation group. There is, however, the additional restriction, based on empirical evidence, that only the two representations of degree (dimension) 1 occur, namely that Ψ must be a symmetrized function Ψ_{SA} , that is, either symmetric (Ψ_S) or antisymmetric (Ψ_A) under any permutation of the particle labels.

On the conceptual side, in classical mechanics similarity or indistinguishability always means essentially the indistinguishability of equal billiard balls. Imagine one billiard ball in your right hand and the other in your left hand. You are blindfolded and somebody else takes the balls out of your hands, then puts them back. If you look at them again you cannot tell whether or not they have been interchanged. However, if you were not blindfolded you could follow their paths and decide which ball was initially in your right hand and which was in your left hand. Or, imagine a situation where the balls are in rapid movement around each other so that you see nothing but a fuzzy cloud of whirling balls. Nevertheless, when you are allowed to use more refined methods of observation you will always be able to follow the paths of the balls individually.

Clearly, in classical physics with its mass points representing the centers of mass of impenetrable bodies, there is no indistinguishability that could not be resolved in principle. In quantum physics this is no longer true. There are situations, associated with wave function overlap, where the observer is unable, in principle, to distinguish the particles, in the sense that he is unable to follow the path of a given particle unmistakably through all processes.

In other words, for him the particles lose their individuality. In the spirit of the realist program this is now conceived, not as any incapability on the part of the observer, but as an objective physical process: the original wavepackets *coalesce* and form a new physical object, a *coalesced system*. And a coalesced system may again coalesce with other similar wavepackets or other coalesced systems.

We conclude that the distinctive mathematical feature of quantum mechanical indistinguishability is just the description of its distinctive conceptual feature. In other words, symmetrization is a mathematical method of coping with the physical process of coalescence: a symmetrized wave function Ψ_{SA} describes a coalesced system.

This fits with the fact that only in the two representations of degree 1 of the permutation group are all particles always treated equally, in that transposition of any two particle labels always multiplies the wave function by the same factor (+1 or -1, respectively; the Young diagrams having a single line or a single column, respectively). A priori the factor could be any complex number of modulus

1 (preservation of the norm). The additional requirement that Ψ form a representation space of the permutation group then restricts this number to either +1 or -1. All this fits with the fact that symmetrization is strictly bound to the *particle* labels and does not work with generalized coordinates such as are permitted in the classical Hamilton formalism. Notice also that in symmetrization we speak of the mathematical operation of exchanging particle labels, not of the physical process of transporting real particles from one place to the other.

A necessary condition for coalescence to set in is that the wavepackets, which are initially separated and independent, at some instant of time overlap in x space:

$$\psi_1(\mathbf{x}, t)\psi_2(\mathbf{x}, t) \neq 0, \quad (16)$$

although the exact degree of the necessary overlap cannot be specified at the present time.

Thus the general symmetrization postulate of conventional quantum mechanics in our interpretation is restricted to apply only to a coalesced system. Consider an electron on the Earth and another in the Andromeda Galaxy immediately after its creation in the decay of a muon. According to the conventional view the wave function describing the system of these two electrons must be symmetrized from the beginning. In our view it must not; symmetrization is to be applied only if the electrons should come together, overlap and coalesce. Certainly, if the wave functions of the two individual electrons do not overlap the properties of the electron on the Earth (and of that in the Andromeda Galaxy) are independent of whether or not the total wave function is symmetrized.⁽¹⁰²⁻¹⁰⁴⁾ Nevertheless, symmetrization will introduce correlations between the electrons, and any such correlations in our interpretation can arise only from coalescence. We shall come back to this point below.

Only similar wavepackets can coalesce. Similarity of wavepackets in this situation means no more and no less than capability of coalescence. We might therefore replace the word “similar” (“identical”) with “coalesceable”. And the criterion for particles to be identical is in our interpretation a criterion for wavepackets to be coalesceable. This is the other reason why we prefer the denomination similar to identical. In the realistic interpretation the wave function represents the particle, and two particles with the same intrinsic quantum numbers but with different wave functions might well be regarded as different although still similar and hence coalesceable.

In addition to the process of coalescence there is the inverse process, the *splitting* of a coalesced system into two or more completely independent systems or wavepackets.

Consider the “exchange terms” or “exchange integrals” that arise from symmetrization in collision processes, in the hydrogen molecule or in para and ortho helium. These terms always go to zero when the distance between the original constituents in some sense becomes large. We infer that coalescence and splitting alternate in those systems and that the exchange term is some measure of the average fraction of time in which the constituents form a coalesced system.

Once coalescence has set in the system is described by a symmetrized configuration space wave function $\Psi_{\text{SA}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N, t)$, that is, either by a symmetrical or by an antisymmetrical function. The mathematical description of the coalesced system in configuration space does not prevent us from visualizing it in ordinary (3+1)-dimensional space: in the Copenhagen interpretation the probability that the position of any one of the N particles is found within d^3x at \mathbf{x} is⁽⁵⁵⁾

$$P = d^3x \sum_{i=1}^N \int |\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{i-1}, \mathbf{x}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_N, t)|^2 \times d^3x_1 d^3x_2 \cdots d^3x_{i-1} d^3x_{i+1} \cdots d^3x_N \quad (17)$$

(where terms with \mathbf{x}_0, d^3x_0 and $\mathbf{x}_{N+1}, d^3x_{N+1}$ have to be omitted). This holds even for dissimilar particles. For similar particles (bosons or fermions) all terms in the sum are equal and we obtain

$$P = \rho(\mathbf{x}, t) d^3x = d^3x N \int |\Psi_{\text{SA}}(\mathbf{x}, \mathbf{x}_2, \dots, \mathbf{x}_N, t)|^2 d^3x_2 \cdots d^3x_N. \quad (18)$$

In our interpretation $\rho(\mathbf{x}, t)$ is a density and is readily visualized. In fact, $\rho(\mathbf{x}, t)$ not only permits a pictorial view of the coalesced system but is also the basis for rigorous calculations of atomic and molecular structure and behavior.⁽¹⁰⁵⁾

Any configuration space function, we recall,^(106–108) can be expanded in terms of a complete set of one-particle functions in ordinary space $\varphi_{r_i}(\mathbf{x}, t)$

$$\Psi_{\text{SA}}(\mathbf{x}_1, \dots, \mathbf{x}_N, t) = \sum_{r_1, \dots, r_N} c(r_1, \dots, r_N, t) \varphi_{r_1}(\mathbf{x}_1, t) \cdots \varphi_{r_N}(\mathbf{x}_N, t). \quad (19)$$

The \mathbf{x}_i may include the spin components. The functions φ_{r_i} are here taken as functions of the time too, as in the Dirac or interaction picture. $c(r_1, \dots, r_N, t)$ is the transformed wave function. In the case of similar wavepackets (“identical particles”) it is completely determined if we specify the number of times each of the arguments r_1, r_2, r_3, \dots occurs in it. These numbers are the occupation numbers n_{r_i} , and the set $|n_1, n_2, n_3, \dots, t\rangle$ is another representation of the wave function $\Psi_{\text{SA}}(\mathbf{x}_1, \dots, \mathbf{x}_N, t)$. The set $|n_1, n_2, n_3, \dots, t\rangle$ is the state (vector, wave function) in the occupation-number, N , or Fock representation.^(108–111) The change in occupation numbers in the course of time, due to interactions, is described by the creation and annihilation operators a^\dagger and a . This works for bosons ($n_{r_i} = 0, \dots, \infty$) and, with minor additions, for fermions ($n_{r_i} = 0, 1$).

The occupation-number representation is well suited for our interpretation because numbering of quanta is not even mentioned in it. The denomination “occupation number” is, however, likely to mislead one to think that there are two different things, the quanta and the states filled with them. In our interpretation these two are the same and therefore we choose a different formulation: the one-particle basis functions are wavepackets, and occupation numbers of 2 or more mean that two

or more wavepackets have *condensed* to form one single wavepacket. Thus boson wavepackets can condense, but fermi packets are excluded from doing so. This is our formulation of the Pauli exclusion principle.

The inverse process is also conceived: *decondensation*. The change in occupation numbers will then be described as alteration, condensation and decondensation of wavepackets. Condensation and decondensation occur only between Bose but not between Fermi packets; alteration occurs with both of them. In Sec. 6 we shall show how these processes lead to a new derivation of the quantum-statistical Bose and Fermi distributions and to the corresponding fluctuations. Condensation, decondensation and alteration are to be distinguished from coalescence and splitting, since they change only the structure of the coalesced system. Coalescence and splitting build and break up coalesced systems.

Coalescence may persist even when parts of the coalesced system have largely separated and are no longer spatially connected, that is, when the regions in space where the density $\rho(\mathbf{x}, t)$ [Formula (18)] is different from zero are spatially disconnected. Experimental examples will be discussed in Sec. 5.3.

Nevertheless, the fact that the disconnected regions are parts of one and the same coalesced system implies specific correlations between them, which means another kind of nonlocality, besides that which stems from the internal structurelessness of the wavepackets themselves. In 1932 Ehrenfest⁽¹¹²⁾ had already indicated that the mere use of a configuration space wave function in quantum mechanics implies some kind of simultaneous action-at-a-distance. Indeed, the conditional probability that particle 1 acts in d^3x_1 about \mathbf{x}_1 (or: that the random variable “position” assumes a value in d^3x_1 about \mathbf{x}_1 , or: that the result of a position measurement is a position about \mathbf{x}_1), if particle 2 acts in d^3x_2 about \mathbf{x}_2 , is

$$P(\mathbf{x}_1|\mathbf{x}_2, t)d^3x_1 = \frac{|\Psi(\mathbf{x}_1, \mathbf{x}_2, t)|^2 d^3x_1 d^3x_2}{d^3x_2 \int |\Psi(\mathbf{x}_1, \mathbf{x}_2, t)|^2 d^3x_1}, \quad (20)$$

and this depends on \mathbf{x}_2 , that is, on the result of a simultaneous action of (observation upon) the second particle. On the other hand, the distance between the particles in ordinary space $|\mathbf{x}_1 - \mathbf{x}_2|$ may be many kilometers long. It is essential that $\Psi(1, 2)$ has an *entangled* form, that is, not the simple product form $\psi(1)\psi(2)$ when written in terms of one-particle functions. With the product form the probability (20) would be independent of \mathbf{x}_2 , and there would be no correlations. Now, symmetrization is a patently entangling procedure and leads to correlations even if the original nonsymmetrical wave function had the product form. Examples of nonlocal or EPR correlations that arise in this way are discussed in Sec. 5 and Appendix C.

One might object that the above kind of nonlocal features in the theory appear only because we are applying nonrelativistic quantum mechanics, where effects can propagate anyway with infinite velocity. We do not think so. Notice that the nonlocal features under discussion are not dynamic effects, which are described by special terms in the (Hamilton, Lagrange etc.) formalism, but appear because a configuration space function is employed which is symmetrized and which depends

on only one time variable t . Much as the correlations stemming from Pauli's exclusion principle are not dynamic but symmetrization effects. After all, the predicted nonlocal effects have been confirmed by experiment (Sec. 5.3), in particular between photons, where there are no dynamic interactions at all. Also, they do not contradict relativity since they do not make superluminal signaling possible (Sec. 5.2).

Whether or not a relativistic theory predicts nonlocal effects depends on the specific structure of that theory. Relativistic theories may be possible that exclude nonlocal correlations, and these might be welcome as a proper treatment of aggregates of dissimilar particles (Sec. 3.3). On the other hand, there is no a priori reason why a relativistic theory should be incompatible with nonlocality. Actually, a relativistic theory that permits nonlocality, in the way here described, is just what is wanted in our interpretation as a proper relativistic treatment of a coalesced system of similar particles.

In our interpretation nonlocal effects may thus arise from two features: (1) the formation of coalesced systems of wavepackets, as accounted for by symmetrized Schrödinger functions in configuration space, and (2) the internal structurelessness of the wavepackets themselves, as introduced as a postulate to account for the "particlelike" behavior in Sec. 2.3. *These two are conceived to be the only sources of nonlocality.*

3.2 Condensed Wavepackets

We want to consider some additional aspects of the condensed wavepackets. They are wavepackets in the technical sense of our interpretation, particularly with the property of internal structurelessness. When several wavepackets condense the sum of the quanta that were previously represented by the individual condensing wavepackets now is represented by the condensed packet; and analogously in decondensation. In other words, the number of quanta is conserved in condensation and decondensation. In Sec. 2.4 we interpreted the normalization of the wave function as meaning that each wavepacket represents one quantum. We now go further and say that each wavepacket, elementary or condensed, represents an *integer number of quanta*, even though Ψ is still normalized to 1 in the current formalism. One might say that a wavepacket *contains* an integer number of quanta but this is a language we wish as a rule to avoid. If one insists in this language one has to use it warily and one should at least not picture the wavepackets as bags containing balls but as bags that contain water in integer multiples of some standard portion.

In an interaction with another packet an N -quantum condensed wavepacket may act with any number of the quanta it represents. Considering the quantization of charge, spin etc. as special manifestations of the existence of quanta we conceive that a condensed packet may act with a fraction of its charge etc., where, however, that fraction is an integer multiple of an elementary charge etc. For example, a condensed packet of two elementary charged-pion packets of charge e each, has the charge $2e$, but it may act with the charge $1e$ only. An N -quantum condensed packet may also undergo up to N interactions simultaneously. In the case

of photon packets this is our view of the effects of Hanbury Brown and Twiss⁽¹¹³⁾ (see also Ref. 114) and of nonlinear optics. Other examples are the photons arising from parametric down-conversion in photon correlation experiments.^(115–118) These photons in our language are elementary or condensed photon wavepackets and together they form a coalesced system.

Finally we want to consider a special property of a condensed wavepacket which we shall use in Sec. 6.5 on quantum-count fluctuations. As the condensed packet arises from the process where some of the one-particle functions in the expansion of Ψ_S (19) become equal we may describe it by means of the product

$$\Psi_S = \varphi(\mathbf{x}_1, s_1, t) \varphi(\mathbf{x}_2, s_2, t) \cdots \varphi(\mathbf{x}_N, s_N, t) \quad (21)$$

where the s_i signify the spin variables, and the same φ is used in all factors. $\varphi(\mathbf{x}, s, t)$ here does not necessarily mean the lowest-energy one-particle state, as it does in Bose-Einstein condensation proper.⁽¹¹⁹⁾ We may mention that in the treatment of laser coherence^(120,121) the wave function of a stationary N -photon state can also be written as a product of the type (21). The condensed packet is thus effectively described by the one function $\varphi(\mathbf{x}, s, t)$ in ordinary space. Consider the expression

$$\begin{aligned} P_1 &= \int_{\mathbf{x}_1 \in D^3} \int_{\mathbf{x}_i \in R^3 \ (i \geq 2)} \cdots \int |\Psi_S(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N, t)|^2 d^3x_1 d^3x_2 \dots d^3x_N \\ &= \int_{\mathbf{x} \in D^3} |\varphi(\mathbf{x}, t)|^2 d^3x. \end{aligned} \quad (22)$$

$\mathbf{x} \in D^3$ means $x_a \leq x \leq x_b, y_a \leq y \leq y_b, z_a \leq z \leq z_b$. R^3 means total space. If necessary \mathbf{x}, D^3 and R^3 are to include the spin variables. For $D^3 = d^3x$ the expression reduces to $P_1 = |\varphi(\mathbf{x}, t)|^2 d^3x$.

In the Copenhagen interpretation expression (22) means the probability that particle 1 of a system of N similar particles is found in the spatial region D^3 , irrespective of where the other $N - 1$ particles are found. In our interpretation it is the probability that wavepacket 1 acts in D^3 , irrespective of where the other $N - 1$ wavepackets act. Actually, the wavepackets are equal, so what we really want is an expression for the probability that any one packet acts in D^3 , irrespective of where the other $N - 1$ packets act, which is $N \cdot P_1$. In the same way the standard formalism gives the probability P_2 that any m wavepackets of the N act in D^3 while the other $N - m$ do not act in D^3

$$\begin{aligned} P_2 &= \binom{N}{m} \int_{\mathbf{x}_1 \in D^3} \cdots \int_{\mathbf{x}_m \in D^3} \int'_{\mathbf{x}_{m+1} \in R^3} \cdots \int'_{\mathbf{x}_N \in R^3} |\Psi_S(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N, t)|^2 \\ &\quad \times d^3x_1 d^3x_2 \cdots d^3x_N \end{aligned} \quad (23)$$

where a prime at the integral sign means that in the integration over the variables \mathbf{x}_i the region D^3 has to be excluded. The combinatorial factor $\binom{N}{m}$ is just the

number of ways m billiard balls (or particle labels) can be chosen from N . In the case of the condensed packet of the product form (21) P_2 reduces to

$$P_2 = \binom{N}{m} \eta^m (1 - \eta)^{N-m} \quad (24)$$

with

$$\eta = \int_{\mathbf{x} \in D^3} |\varphi(\mathbf{x}, t)|^2 d^3x. \quad (25)$$

This is the binomial distribution. It is just the probability that N independent trials with probabilities η for success and $1 - \eta$ for failure result in m successes and $N - m$ failures. We may thus say that P_2 is the probability that the condensed packet acts with m of its N quanta in D^3 .

In Sec. 6.2 we shall see that there are many other objects, conceived long ago in radiation theory, that are very similar to the condensed packets and may be taken to be other aspects of them.

3.3 Dissimilar Wavepackets: Aggregates

We come now to the quite different case of dissimilar wavepackets (“non-identical particles”). In the Copenhagen interpretation where the wave function is the mathematical description, not of a physical object, but merely of the observed behavior of a physical object, it does not appear to make much difference whether the object concerned is an electron, an atom, a macroscopic body, the universe, or a fictitious point like the center of mass. This is different in our interpretation where the wave function is the mathematical description or representation of the individual physical object itself. And the physical object that is so represented in our interpretation is only a wavepacket, [mathematically represented by $\psi(\mathbf{x}, t)$], or a coalesced system of similar wavepackets [mathematically represented by the symmetrized configuration space wave function Ψ_{SA}].

A system of dissimilar elementary particles like an atom, a macroscopic body, or the universe in our interpretation is neither a wavepacket nor a coalesced system but an *aggregate* of wavepackets. Generally, an aggregate is a set of distinct wavepackets, which may or may not dynamically interact with one another. The hydrogen atom, for example, is an aggregate of a proton wavepacket and an electron wavepacket. A system of two electrons, on the other hand, may be a coalesced system or an aggregate, depending on whether the two electron wavepackets have coalesced or not. In those aggregates that contain similar wavepackets some wavepackets may coalesce and the coalesced system may split again. Except for this complication, in our conception of an aggregate we essentially replace the mass points of the classical Newtonian description by wavepackets. If all wavepackets and coalesced systems of an aggregate were mathematically specified the description would be complete; much as the classical description is complete if the positions and momenta of all the mass points are specified.

The collapse or reduction process does not apply to the aggregate as a whole, only the wavepackets and the coalesced systems that constitute the aggregate may be reduced. Ludwig^(122,123) seems to have drawn similar conclusions although through different approaches.

Now, $\psi(\mathbf{x}, t)$ in our interpretation is an elementary or condensed wavepacket, and Ψ_{SA} is a coalesced system of similar wavepackets, and we may ask what is the non-symmetrized configuration space wave function $\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N, t)$? This function in our interpretation is no more than a convenient calculational device that serves to compute some properties and effects of systems of dissimilar particles. It may be compared, for example, with the partition function in statistical mechanics. Table I summarizes the correspondence between mathematical symbols and physical objects.

Table I: Mathematical Symbols and Their Corresponding Physical Objects in the Copenhagen and in the Realistic Interpretation

However, we do not think that all conclusions drawn from the function Ψ correspond to physical reality. As we have already pointed out in Sec. 3.1 the mere use of a configuration space wave function in quantum mechanics already implies

some kind of simultaneous action-at-a-distance and nonlocal correlations, provided Ψ happens to be entangled and the usual relation between wave function and probability holds. It is a basic conception of our interpretation that for systems of dissimilar wavepackets these correlations do not represent physical reality. In our interpretation the nonlocal correlations can only arise from the special physical process of coalescence between similar wavepackets. Thus, the predictions concerning nonlocal correlations implied by entangled configuration space wave functions are to be taken seriously only in so far as they describe coalesced systems, that is, if the entanglement results from symmetrization.

What would then be a proper mathematical description of a system of N dissimilar elementary particles in our interpretation? We do not have the answer and can only present some considerations on this problem in the remainder of this section.

Of course we do not deny that for instance the usual treatment of the hydrogen atom, which sets out from 6-dimensional two-body configuration space, leads to correct wavelengths and intensities of the emitted spectral lines. The problem is that it also leads to nonlocal correlations.

The basic physical elements in systems of dissimilar particles being wavepackets in ordinary space one would like to have a description in terms of N $(3 + 1)$ -dimensional functions $\psi_i(\mathbf{x}_i, t)$, $i = 1, 2, \dots, N$, one for each particle. These functions might be determined by a system of coupled equations, in analogy to the functions $\mathbf{x}_i(t)$ in a system of Newtonian point particles. An example (without guarantee of being physically right) is given by the formulas (27) below. The probability that particle i produces an effect at \mathbf{x}_i might be given by $|\psi_i(\mathbf{x}_i, t)|^2 d^3x_i$ [not by a term of the sum (17), as usual], whatever configuration space functions Ψ might be constructed from the ψ_i 's for other purposes. The particles may of course still depend on one another because the determining equations are coupled, but this does not necessarily imply nonlocality. In standard nonrelativistic quantum mechanics such a description does not exist, and I would consider this place as one of those, alluded to in Sec. 1.2, where the traditional formalism is expected to be changed in future developments. In support of this attitude I want to adduce the following observations.

Einstein writes:

Of the new attempts to obtain a deeper formulation of the quantum laws I like the one of Schrödinger most. If there introduced wave fields could only be transplanted from the n -dimensional coordinate space into the 3 resp. 4 dimensional one!⁽¹²⁴⁾

In the case of similar particles this does not appear so bad in the new interpretation, because here the N -particle configuration space means a special physical system, but in the case of dissimilar particles Einstein's remark still carries its full weight. Actually, two such descriptions of systems of dissimilar particles already exist: in canonical quantum field theory and in Feynman's space-time approach to quantum electrodynamics. There, each particle is described by one and only

one particular spacetime function $\psi_i(\mathbf{x}, t)$. However, use is made there of the procedure of canonical quantization, or equivalent procedures, and these lead to the well known difficulties. Therefore we cannot use these descriptions to support our concept of an aggregate.

Recently, Barut^(83,125,126) has taken up the problem of individual wave functions in 3-dimensional space within his approach to electrodynamics with self-interaction but without canonical quantization. The basic idea, formulated for a system of a proton and an electron, is the following. One starts in the usual way with a Lagrangian density \mathcal{L} for two classical Dirac fields $\psi_1(x)$ and $\psi_2(x)$ and an electromagnetic field produced by the currents of the fields⁽¹²⁷⁾

$$\mathcal{L} = \sum_{j=1}^2 \{ \bar{\psi}_j(x)(i\gamma^\mu \partial_\mu - m_j)\psi_j(x) - e_j \bar{\psi}_j(x)\gamma^\mu \psi_j(x)A_\mu(x) \} - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \quad (26)$$

$$F_{\mu\nu}(x) = \partial_\mu A_\nu(x) - \partial_\nu A_\mu(x).$$

The variation of the action integral $W = \int \mathcal{L}d^4x$ with respect to ψ_1, ψ_2 and the A_μ , respectively, leads to the well known coupled differential equations

$$\begin{aligned} (i\gamma^\mu \partial_\mu - m_1)\psi_1 &= e_1 \gamma^\mu A_\mu \psi_1 \\ (i\gamma^\mu \partial_\mu - m_2)\psi_2 &= e_2 \gamma^\mu A_\mu \psi_2 \\ F^{\mu\nu}{}_{,\nu} &= -e_1 \bar{\psi}_1 \gamma^\mu \psi_1 - e_2 \bar{\psi}_2 \gamma^\mu \psi_2. \end{aligned} \quad (27)$$

Barut found out that W can be written as a function of the product $\Phi(x_1, x_2) = \psi_1(x_1)\psi_2(x_2)$, and that the variation of W with respect to this Φ leads to a linear two-body Schrödinger equation for Φ in configuration space. Since this Φ is constructed from individual 3-dimensional functions, one for each particle, Barut concludes:

A further objection has always been the use of the configuration space wave function $\psi(x_1, x_2, \dots, x_n)$ in many-body problems. It turns out that one can discuss the many-body problem in terms of the individual wave functions $\psi_1(x), \psi_2(x), \dots$ and their individual currents. ... Thus, the use of configuration space wave function does in no way exclude that the individual particles have their own current distributions.⁽¹²⁵⁾

The linear equation obtained for Φ in configuration space permits a solution of the form $\Phi = \varphi_1(\mathbf{R})\varphi_2(\mathbf{r})$, where \mathbf{R} is the coordinate of the center of mass and \mathbf{r} that of a fictitious particle with reduced mass. A solution that would fit with the wavepacket picture of an aggregate should have the form $\Phi = F[\psi_1(x_1), \psi_2(x_2)]$, where $\psi_1(x)$ describes the electron and $\psi_2(x)$ the proton. Such a solution is not given by Barut, but it may perhaps still be found. Alternatively, one might consider directly the system of coupled Eqs. (27) for the functions ψ_1 and ψ_2 . This would be interesting also from the following point of view. In the Copenhagen quantum mechanics the treatment of the hydrogen atom starts with a two-body function

in configuration space. The problem is then reduced to two one-body problems in 3-dimensional space, where the one body is the center of mass and the other a fictitious particle with reduced mass. Eigenfunctions belonging to the center of mass are plane waves and eigenfunctions belonging to the fictitious particle are the Coulomb functions ψ_{nlm} . Both types of functions describe features of physical reality: the ψ_{nlm} describe the spectral lines etc. and the plane waves describe the diffraction patterns in experiments with atomic beams.^(128,129) In these experiments, is it the center of mass that produces the spots that make up the diffraction pattern? Physicists would certainly prefer saying that it is the electron and the proton, but how exactly this is brought about could only be resolved via a knowledge of $\psi_1(x)$ and $\psi_2(x)$.

3.4 Summary of Principles and Postulates

The preceding section concluded the presentation of the basic features of our interpretation. In the following sections we shall apply them to specific problems. Let us pause here for a short summary. This is not meant as an axiomatization, but only as a list of the most salient features that are different from those of the Copenhagen interpretation. A somewhat enlarged summary is Ref. 130, and a preliminary account was given in Ref. 131.

(1) The function ψ (in the Schrödinger picture, say) is not a probability amplitude but an objective physical field, comparable to the classical electromagnetic field (Sec. 1.2).

(2) Any particular normalized ψ function which in the Copenhagen interpretation describes the observed behavior of an elementary particle is in our interpretation the individual particle itself. An elementary particle is defined as a particle listed in the Particle Properties Tables. The particle thus is an extended object, of size $\Delta x = \langle (x - \langle x \rangle)^2 \rangle^{1/2}$ say, but the size is not fixed: it varies in time according to the variation of $\psi(\mathbf{x}, t)$.

An elementary particle thus described is called a wavepacket, or, more specifically, an elementary wavepacket. Another object that is also a wavepacket is the condensed wavepacket of point 6 below. Wavepackets, elementary or condensed, are the basic objects in our interpretation.

The reduction or collapse of the wavepacket, which in the Copenhagen interpretation occurs in and only in a measurement, in our interpretation is an objective physical process which occurs in special physical situations independent of whether these are measurements or not (Sec. 1.3).

(3) The expression $|\psi(\mathbf{x}, t)|^2 d^3x$ is not defined as a probability. It is, in certain circumstances, numerically equal to a probability, but to one that is different from the Copenhagen position probability and may be called an action probability: it is the probability that the wavepacket, when placed into an appropriate physical environment, induces an effect (observed or not) within the volume d^3x about \mathbf{x} and in a small but otherwise arbitrary time interval Δt about t .

This interpretation of $|\psi|^2 d^3x$ can be justified by a calculation that is based on the transition-probability formulas of the standard formalism. In these formulas the probability refers only to the occurrence of the transition, not to the nature

of the thing that makes the transition. The circumstances in which this interpretation turns out to be possible include those where the physical environment of the wavepacket, in the language of conventional quantum mechanics, allows the position of the object to be directly measured. Examples are bubble chambers and photographic emulsions. The exact value of the time interval Δt drops out because the environment is meant to be homogeneous, time independent and such that the wavepacket with certainty induces an effect anywhere in it *during* Δt . Of course, Δt must be short enough so that $|\psi|^2$ can be considered constant in it. If the actual environment meets these requirements only partly the raw data must be corrected for this, and $|\psi|^2 d^3x$ refers to the corrected data.

The interpretation in the case of quantities other than position is analogous. Since we may say that the wavepacket with probability $|\psi(\mathbf{x}, t)|^2 d^3x$ induces an effect at \mathbf{x} (“in d^3x about \mathbf{x} ”), or acts at \mathbf{x} in position space, we may say that if the wavepacket in some “ A space” is $\varphi(a, t)$, the expression $|\varphi(a, t)|^2 da$ gives the probability that the wavepacket acts at a in A space. This also applies to discrete eigenvalues where “at a ” then means “at the point a ” rather than “in da about a ” (Sec. 2.2).

(4) A wavepacket has no internal structure. The wavepacket is an elementary region of space and must not be subdivided into smaller parts. In other words: there are no interactions between parts of one and the same wavepacket.

This internal structurelessness implies that the wavepacket in a sense behaves like a rigid body of Newtonian mechanics, and this means some kind of nonlocality. We speak of internal structurelessness and not just of structurelessness because the absence of structure comes into play only when the intrinsic features of the interaction as such are considered, not when the probability of its occurrence is calculated. In this calculation the full $\psi(\mathbf{x}, t)$ is used, defined for every point in the wavepacket (Sec. 2.3).

(5) Systems of similar particles in traditional quantum mechanics are described by symmetrized wave functions in configuration space $\Psi_{SA}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N, t)$. This expresses the fact that for the observer the particles are indistinguishable and lose their individuality. This loss is now conceived to be a physical process: if the similar wavepackets at some instant of time overlap in space they may coalesce and form a new physical object: a coalesced system. A symmetrized (symmetric or antisymmetric) N -particle configuration space wave function thus serves only to describe a special N -particle system.

Coalescence leads to some kind of action-at-a-distance between the coalesced wavepackets and may give rise to nonlocal effects. Coalescence and internal structurelessness are conceived to be the only sources of nonlocality in nature.

In addition to coalescence there is the inverse process of splitting of a coalesced system into several distinct coalesced systems or wavepackets (Sec. 3.1).

(6) There is coalescence both between similar Bose packets and between similar Fermi packets. The difference is that Bose packets in addition can condense, but Fermi packets cannot. Condensation is a process where several of the one-particle functions that build up the configuration space function Ψ_S become equal. There

is also the inverse process: decondensation. The number of quanta is conserved in condensation and decondensation. The condensed wavepacket represents an integer number of quanta (not just one) but in every other respect is a wavepacket in the technical sense of the interpretation, particularly with the property of internal structurelessness. Condensation, decondensation and alteration are our interpretation of the change in occupation numbers when Schrödinger's configuration space formalism with symmetrized functions is cast into the form of the occupation-number representation (Sec. 3.1).

(7) A system of dissimilar elementary particles like an atom, a macroscopic body, or the universe, is neither a wavepacket nor a coalesced system, but an aggregate of wavepackets. We essentially replace the Newtonian mass points of the classical description by wavepackets. Once all wavepackets of such an aggregate are mathematically specified at all times (details unknown at present) the description is complete. The collapse or reduction process does not apply to the aggregate as a whole. Only the individual constituent wavepackets and coalesced systems may be reduced.

The usual description of systems of dissimilar particles by means of a non-symmetrized configuration space wave function $\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N, t)$ is just a convenient calculational device that serves to compute some properties and effects of these systems. It is considered misleading when it comes to the prediction of action-at-a-distance (Sec. 3.3).

4. MEASUREMENTS AND TRANSITIONS

4.1 The Copenhagen Measurement Postulates

As a first application, in this section the problem of measurement in quantum mechanics will be discussed. This problem is indicative of the basic difficulties of the Copenhagen interpretation.

The basic postulates concerning measurements in the Copenhagen quantum mechanics may be expressed in the following way:

1) Let $\psi(\mathbf{x}, t)$ be the normalized wave function in x space. The expression $|\psi(\mathbf{x}, t)|^2 d^3x$ is then the probability that the result of a position measurement at time t lies in the interval d^3x . Or, more generally, let $c(n, t) = \int \psi(\mathbf{x}, t) \varphi_n^*(\mathbf{x}, t) d^3x$ be the wave function in A space where the quantity A is represented by a self-adjointed operator with discrete nondegenerate eigenvalues a_n and eigenfunctions φ_n . Then the expression $|c(n, t)|^2$ is the probability that the result of a measurement of A is the value a_n .

2) The only values that the quantity A may assume in the measurement are those of the eigenvalue spectrum of the operator associated with A . This is the usual, somewhat sloppy formulation. It is empty in the case of continuous eigenvalues which cover the whole real axis. In the case of a discrete spectrum we may say, more precisely, that the only values that can be found will lie in arbitrarily small intervals about the discrete eigenvalues. Notice that if the result of the measurement is the eigenvalue a_n , it cannot be said that the considered system before

the measurement had the value a_n , rather the value a_n comes into existence by the act of measurement, and only immediately after the measurement the system is said to have the value a_n .

3) The third postulate introduces a strong assumption regarding the form of the wavepacket after the measurement. It has been worked out very clearly by von Neumann⁽¹³²⁾ and is therefore often called von Neumann's axiom. It assumes that immediately after the measurement the wavepacket is a superposition of eigenfunctions of the respective operator which belong to the interval of eigenvalues specified by the measurement.⁽¹³³⁾ This statement covers the case of continuous as well as discrete eigenvalues. In the case of continuous eigenvalues the measurement interval can be identified with the above interval d^3x . In the extreme case of a discrete spectrum, a nondegenerate eigenvalue and a sufficient measurement accuracy, the wavepacket immediately after the measurement will be a completely specified eigenfunction of the respective operator.

The proviso "immediately after the measurement" emphasizes the fact that if we wait too long, the wavepacket has already changed in accordance with the time dependent Schrödinger equation.

All this refers to "ideal" measurements, i.e. where any change in the measured wavepacket due to effects other than those mentioned above can be neglected. For example, the components of the momentum of a charged particle are changed when they are measured by means of the deflection of the particle in a magnetic field. This is a calculable effect. It is assumed that such effects can always be completely reduced or compensated.

All this also applies only to specific quantum measurements by which we mean those where the wavepacket nature of the elementary particles cannot be neglected, that is, where a particular quantity is to be measured with an error interval that is smaller than, or of the order of, the corresponding width of the wavepacket. In other measurements the above postulates lose their relevance and the elementary particles may be treated like classical objects, for example when the electron's e/m ratio is measured. This is in accordance with our definition of the classical limit of quantum mechanics given in Sec. 2.5.

According to the postulate 3 the measurement effects a reduction of the initial wavepacket: the initial packet may be expanded mathematically in a series (or an integral) of eigenfunctions φ_n of the self-adjointed operator belonging to the measured quantity

$$\psi = \sum_{n=-\infty}^{+\infty} c(n) \varphi_n \quad (28)$$

where n ranges over all values that specify the complete set of eigenfunctions φ_n , symbolized in (28) by writing $n = -\infty \dots +\infty$. The measurement reduces the wavepacket to only a part of the sum

$$\psi = \sum_{n=-\infty}^{+\infty} c(n) \varphi_n \rightarrow \sum_{n=n_1}^{n_2} c(n) \varphi_n. \quad (29)$$

That is, it narrows down the range of values of n . In the extreme case reduction leads to one single discrete eigenfunction φ_0

$$\psi = \sum_{n=-\infty}^{+\infty} c(n) \varphi_n \rightarrow c(0) \varphi_0, \quad (30)$$

where the normalization of φ has to be re-adjusted so that $\int_{-\infty}^{+\infty} |c(0)\varphi_0|^2 d^3x = 1$. The reduction is therefore also called a collapse, preparation, filtering, or projection of the wavepacket. It clearly means a nonlinear evolution of the wave function: if φ_m and φ_n are two normalized eigenfunctions belonging to different nondegenerate eigenvalues reduction of φ_m leads to φ_m , reduction of φ_n to φ_n , but reduction of $(\varphi_m + \varphi_n)/\sqrt{2}$ does not lead to $(\varphi_m + \varphi_n)/\sqrt{2}$ but to either φ_m or φ_n (Tausk⁽¹³⁴⁾). It is thus a change in the wavepacket that is not described by the Schrödinger equation.

Thus, a wavepacket may vary with time in two ways: sometimes deterministically (as determined by the Schrödinger equation) and at other times with a random element in its behavior (in a measurement). The question is, What feature of the measurement makes the measurement interaction so different from the other one? This is the “measurement problem” of traditional quantum mechanics. In the most orthodox version of the Copenhagen interpretation the reduction occurs when the observer becomes conscious of the result.

The realistic interpretation does not need the above postulates.

Regarding postulate 1, we have already explained in Sec. 2.2 that the expression $|\psi(\mathbf{x}, t)|^2 d^3x$ can be taken to be numerically equal to the probability that the wavepacket $\psi(\mathbf{x}, t)$ induces an effect within d^3x and that from this we cannot, without further steps, draw conclusions regarding the properties of the wavepacket. The derivation of $|\psi(\mathbf{x}, t)|^2 d^3x$ as the probability of inducing an effect will be discussed below, in Secs. 4.2 and 4.3.

Regarding postulate 2, in a realistic formulation a measurement measures what already exists. This implies that many operations which in the Copenhagen interpretation are called measurements in a realistic interpretation are not. Thus, neither is the observation of the position of the black spot produced by an electron in a photographic emulsion a measurement of the electron’s position, as we have seen in Secs. 2.1 and 2.2, nor is the registration of an atomic electron in the spin-up path of a Stern-Gerlach apparatus a measurement of the electron’s spin component, as we shall see in Sec. 4.4. The above operations may indeed be used for a measurement of properties that already exist, but not without further steps. In any case an ensemble of wavepackets must be given that are equal with respect to the considered property, as we have already seen in Sec. 2.2 and in the sugar-cube example in Sec. 2.5.

Regarding postulate 3, we emphasize that von Neumann’s axiom does not just mean that the reduction process occurs, but that it occurs in and only in a measurement or observation. We cannot do away with the reduction process, but we do away with the assertion that it occurs only in a measurement or observation. In

the realistic interpretation the reduction processes are objective physical processes that occur in some physical situations whether these involve a measurement or not. The problem remains, but it assumes a different form: what feature of the physical environment makes the reduction process, rather than the deterministic evolution, occur? One example is the Stern-Gerlach apparatus in Sec. 4.4, and other examples are given in Sec. 4.6.

4.2 Derivation of $|\psi|^2 d^3x$ as an action probability

As explained in Sec. 2.2 the expression

$$P_1 = |\psi(\mathbf{x}, t)|^2 d^3x \quad (31)$$

in our interpretation is numerically equal to the probability that the wavepacket $\psi(\mathbf{x}, t)$, when placed into an appropriate physical environment, induces an effect within d^3x during a small but otherwise arbitrary time interval Δt about t . This probability was called an action probability. We did not prove our assertion in Sec. 2.2, but only quoted the results. In the present section and the next we fill this gap.

For our demonstration we consider the special situation of an electron wavepacket that passes through a screen made of a thin layer of photographic emulsion and produces a black spot at a certain position \mathbf{x} . The emulsion layer also consists of wavepackets (electrons, protons and neutrons), and a black spot of silver grains is the result of a cascade of reactions in the emulsion initiated by the interaction of the incoming electron wavepacket with one of the emulsion packets. We consider a situation where the incoming wavepacket is much larger than the emulsion packets (atoms). Packets as long as 10^4 km may indeed exist in laser beams with large coherence length.⁽¹³⁵⁾ The coherence length of a beam is a lower bound for the length of the wavepackets of which it is composed (Appendix A). Electron packets usually are much smaller than photon packets. The largest coherence length of electron beams so far observed is of the order of 10^{-2} cm (Ref. 136), which is, nevertheless, still much larger than an atomic diameter or even a black spot of silver grains (10^{-7} m). As the initiating interaction of the incoming electron wavepacket we consider its interaction with an atom of the emulsion located within d^3x about \mathbf{x} , such that the atom is ionized. We also suppose that d^3x is chosen larger than the final black spot. Thus, the probability of obtaining a black spot in d^3x about \mathbf{x} is proportional to the probability that the incoming electron interacts with an atom in d^3x , that is, that the electron wavepacket makes a transition from its initial form ψ_i to some final form ψ_f , simultaneously with an atom in d^3x making a transition from the ground state Φ_0 to some state Φ_n . The probability of this transition is the basis of our consideration.

We start with the transition probability (1) written in a Schrödinger representation where the position variables are diagonal

$$W = \left| \int_{-\infty}^{+\infty} \psi_f^*(\mathbf{x}, t) U(t, t_0) \psi_i(\mathbf{x}, t_0) d^3x \right|^2 \quad (32)$$

where $U(t, t_0)$ is the evolution operator (propagator) defined by the relation

$$\psi(\mathbf{x}, t) = U(t, t_0) \psi(\mathbf{x}, t_0), \quad (33)$$

and which for conservative systems can be written as $U(t, t_0) = \exp[-i(t-t_0)H/\hbar]$, with H being the Hamilton operator of the total system. In (32) W can be considered as the probability that the wavepacket changes its form from ψ_i to ψ_f during the interval $\Delta t = t - t_0$. Our aim is to get from W in (32) to P_1 in (31), both formulas being interpreted in the new sense.

In evaluating formula (32) perturbation theory is usually applied. The total Hamiltonian is written as $H = H^{(0)} + V$, V is taken to be small and is considered not to depend on time explicitly. Using an expression like (33) with H replaced by $H^{(0)}$ in $U(t, t_0)$ in first nontrivial order of perturbation theory one obtains

$$W \approx \frac{1}{\hbar^2} \left| \int_{t_0}^t \int_{-\infty}^{+\infty} \psi_f^*(\mathbf{x}, \tau) V(\mathbf{x}) \psi_i(\mathbf{x}, \tau) d^3x d\tau \right|^2 \quad (34)$$

where $\psi_i(\mathbf{x}, \tau)$ and $\psi_f(\mathbf{x}, \tau)$ are now the initial and final forms of the packet as they would appear in time τ if they were influenced only by $H^{(0)}$, according to the Dirac or interaction picture of temporal development.

Consider now the specific situation of the electron when it passes through the emulsion screen and ionizes an atom. The meaning of the transition probability in this situation is amplified : ψ_i and ψ_f are now to be regarded as describing all the initial and all the final wavepackets, respectively, that participate in a transition process. Thus, in our case $\psi_i(\mathbf{x}, \tau)$ in formula (34) will now mean the product of the initial electron function $\psi_i(\mathbf{x}_1, \tau)$ with the initial atom function $\Phi_0(\mathbf{x}_2, \tau)$, and ψ_f will mean $\psi_f(\mathbf{x}_1, \tau)\Phi_n(\mathbf{x}_2, \tau)$. The atom, for simplicity of the argument, may be written as a function of only one coordinate \mathbf{x}_2 . In reality it is a function of more coordinates, one for each electron, proton and neutron, if we neglect that some particles are similar and may coalesce. Formula (34) thus changes into⁽¹³⁷⁾

$$W \approx \frac{1}{\hbar^2} \times \left| \int_{t_0}^t \int_{-\infty}^{+\infty} \psi_f^*(\mathbf{x}_1, \tau) \Phi_n^*(\mathbf{x}_2, \tau) V(\mathbf{x}_1 - \mathbf{x}_2) \psi_i(\mathbf{x}_1, \tau) \Phi_0(\mathbf{x}_2, \tau) d^3x_1 d^3x_2 d\tau \right|^2. \quad (35)$$

Now, in the interactions usually invoked for position measurements the interaction potential $V(\mathbf{x}_1 - \mathbf{x}_2)$ is a short-range potential. This is basically correct for the strong and the weak interactions. The electromagnetic (Coulomb) interactions actually have a long range; but due to the fact that the atoms are neutral and randomly oriented in space (isotropic medium) the range is effectively short. Therefore we may write $V(\mathbf{x}_1 - \mathbf{x}_2) = V\delta(\mathbf{x}_1 - \mathbf{x}_2)$ so that

$$W \approx \frac{1}{\hbar^2} \left| \int_{t_0}^t \int_{-\infty}^{+\infty} \psi_f^*(\mathbf{x}_2, \tau) \Phi_n^*(\mathbf{x}_2, \tau) V \psi_i(\mathbf{x}_2, \tau) \Phi_0(\mathbf{x}_2, \tau) d^3x_2 d\tau \right|^2.$$

Using the mean-value theorem of calculus we extract ψ_i from the integral and write

$$W \approx |\psi_i(\mathbf{x}_3, \tau_3)|^2 \frac{1}{\hbar^2} \left| \int_{t_0}^t \int_{-\infty}^{+\infty} \psi_f^*(\mathbf{x}_2, \tau) \Phi_n^*(\mathbf{x}_2, \tau) V \Phi_0(\mathbf{x}_2, \tau) d^3x_2 d\tau \right|^2 \quad (36)$$

where \mathbf{x}_3 and τ_3 are fixed but unknown values anywhere in $-\infty \leq \mathbf{x}_3 \leq +\infty$ and $t_0 \leq \tau_3 \leq t$. The range in which \mathbf{x}_3 must lie can be narrowed down by observing that the atom is concentrated in a finite region about its center \mathbf{x}_0 so that the integral gets its main contribution from that region. Furthermore, because the atom has been assumed to be very small compared with the incoming wave packet, ψ_i at any moment has almost the same value over the whole atomic region, and we may practically identify \mathbf{x}_3 with \mathbf{x}_0 in (36). Thus we arrive at

$$W = \kappa_1(t, t_0) |\psi_i(\mathbf{x}_0, \tau_3)|^2 \quad (37)$$

with

$$\kappa_1(t, t_0) \approx \frac{1}{\hbar^2} \left| \int_{t_0}^t \int_{-\infty}^{+\infty} \psi_f^*(\mathbf{x}_2, \tau) \Phi_n^*(\mathbf{x}_2, \tau) V \Phi_0(\mathbf{x}_2, \tau) d^3x_2 d\tau \right|^2 \quad (38)$$

which is already proportional to $|\psi|^2$.

Formula (37) was derived by using only first order perturbation theory. Even if the derivation would not work for the higher orders the first order alone is known to give a very accurate approximation, so that the above result is valid at least to the same degree of accuracy. As explained in Sec. 2.2, an approximate result is sufficient for maintaining our interpretation. The essential conditions met in the derivation so far are that the wavepacket $\psi(\mathbf{x}, t)$ is large compared to the registering wavepackets which constitute the measuring apparatus, and that the range of the interaction potential between ψ and the registering packets is short.

In order to arrive at P_1 of (31) we have to bring in the factor d^3x . For this we notice that so far only the interaction of the incoming wavepacket ψ with one atom at \mathbf{x}_0 has been considered. Now let there be the number $\rho(\mathbf{x}_0)d^3x$ of such atoms within the volume d^3x . The quantity $\rho(\mathbf{x}_0)$ is the (time independent) density of atoms about \mathbf{x}_0 . Following the usual procedure we assume that each atom acts as if it were alone; then the probability of an interaction within d^3x is proportional to the number of atoms in d^3x , hence to d^3x . We may thus write this probability as (using \mathbf{x} instead of \mathbf{x}_0)

$$W_1 = \kappa(\mathbf{x}, t, t_0) |\psi_i(\mathbf{x}, \tau_3)|^2 d^3x \quad (39)$$

where $\kappa(\mathbf{x}, t, t_0)$ is now the product $\rho(\mathbf{x}) \kappa_1(t, t_0)$, with $\kappa_1(t, t_0)$ from (38).

Actually, the final forms of the wavepackets are not relevant; they should only differ from the initial forms to the extent that changes have been caused by the production of black spots. We thus should replace the right-hand sides of the transition

formulas [e.g. (36)] by sums over many such right-hand sides, each referring to a different possible final form of the electronic and atomic packets. This corresponds to the total, rather than differential, cross section for black-spot production. Nevertheless, the main result, Eq. (37), would remain unaffected in form because our reasoning did not rely on any particular form of the final packets; hence we could extract $|\psi_i|^2$ from every term of the sum. This amounts to replacing the proportionality factors κ_1 and κ in (37) to (39) by a sum of expressions like that in (38). Therefore we shall assume that $\kappa(\mathbf{x}, t, t_0)$ is such a sum. $\kappa(\mathbf{x}, t, t_0)$ depends on the surroundings of the incoming wavepacket in the measuring apparatus, namely on the density and nature of the atoms with which the incoming packet is to interact, as well as on the admitted set of final forms for ψ_f and Φ_n . The physical situations usually employed for testing the localization probability formulas correspond to a situation where the response of the measuring apparatus does not depend on position or on time: one uses a static and homogeneous emulsion. This means that κ does not depend on \mathbf{x} . Even if there is some inhomogeneity or time dependence in the measuring device it can usually be corrected for, and it is the corrected data to which the formulas refer. The independence of time of the measuring device implies that κ only depends on the difference $\Delta t = t - t_0$. Thus we may write (dropping also the index i)

$$W_1 = \kappa(\Delta t) |\psi(\mathbf{x}, \tau_3)|^2 d^3x \quad (40)$$

as the probability of the initial wavepacket inducing a black spot during Δt in d^3x .

4.3 Continuation: The Time Variable

We continue the derivation of $|\psi|^2 d^3x$ as an action probability by studying the role of the time variable in the probability formulas. In particular, there is the problem that the probability (31)

$$P_1 = |\psi(\mathbf{x}, t)|^2 d^3x \quad (31)$$

in the Copenhagen interpretation means the probability of obtaining a certain result in a measurement *at time* t , whereas the probability (40)

$$W_1 = \kappa(\Delta t) |\psi(\mathbf{x}, \tau_3)|^2 d^3x \quad (40)$$

means the probability of a transition occurring *during the interval* Δt about τ_3 .

Why does the size of Δt not appear in the Copenhagen interpretation of the formula for P_1 ? Because of the nature of the question that is asked there: the measurement to which P_1 refers is a measurement of a property of a physical system, and the length of the time interval needed to detect that property is irrelevant; provided of course that it is short enough to consider that property as constant in it.

On the other hand, W_1 in (40) does not refer to a property of a system but to the occurrence of an event, and we draw conclusions regarding the properties of the system only through the occurrence of the events. Now, in order thus to

measure a property within Δt , the events have to occur within Δt , the same Δt as used above for P_1 . The required smallness of Δt allows us also to put τ_3 of (40) equal to t of (31). Thus the integral

$$I_1 = \kappa(\Delta t) \int_{-\infty}^{+\infty} |\psi(\mathbf{x}, t)|^2 d^3x \quad (41)$$

is equal to the probability that the wavepacket ψ induces an effect anywhere in space during the interval Δt . We may write W_1 as a product $W_1 = I_1 \times P_1$:

$$\underbrace{\kappa(\Delta t) |\psi(\mathbf{x}, t)|^2 d^3x}_{W_1} = \underbrace{\kappa(\Delta t) \int_{-\infty}^{+\infty} |\psi(\mathbf{x}, t)|^2 d^3x}_{I_1} \times \underbrace{\frac{|\psi(\mathbf{x}, t)|^2 d^3x}{\int_{-\infty}^{+\infty} |\psi(\mathbf{x}, t)|^2 d^3x}}_{P_1} \quad (42)$$

and we would arrive at $W_1 = P_1$ if $I_1 = 1$, i.e. if the wavepacket ψ induces with certainty an effect anywhere in space during Δt . This is essentially the requirement that the measuring apparatus should have a 100% overall efficiency for detection within Δt . This we shall assume because it corresponds to what is assumed in Copenhagen quantum mechanics (Sec. 2.2). If the actual apparatus has less than 100% efficiency the raw data can be corrected for this. In this correction the value of Δt of course enters since the efficiency of the apparatus for causing the wavepacket to induce an effect depends on the available time interval. Formula (31) refers to the corrected data, hence here the value of Δt no longer appears and we have arrived at our desired result that P_1 is equal to the probability that the wavepacket $\psi(\mathbf{x}, t)$ induces an effect within d^3x during a short but otherwise arbitrary time interval about t .

We want to add a few comments. The Copenhagen interpretation of the expression $|\psi(\mathbf{x}, t)|^2 d^3x$ is the localization probability within d^3x at time t , and not: *within the time interval dt* . Thus space and time are treated in different ways. This is a general feature of traditional quantum mechanics. Time is treated as a parameter in the ψ function. There is a position operator but there is no time operator (although in the momentum-energy space of the Fourier transforms there is both a momentum and an energy operator). This does not mean that the formalism of the theory cannot be put into a covariant form. The Dirac and the Klein-Gordon equations are manifestly covariant equations, and in them space and time appear on an equal footing. It is because of the different types of questions that are asked about space and time respectively that space and time receive different treatment. This different treatment is one of the major obstacles to a reconciliation of quantum theory and relativity theory.⁽¹³⁸⁾

Our requirement that the wavepacket induces with certainty an effect anywhere in space during Δt does not imply that ψ is normalized, $\int |\psi|^2 d^3x = 1$. Rather it means, according to (41), that $I_1 = \kappa(\Delta t) \int |\psi|^2 d^3x = 1$. The additional postulate (Sec. 2.4) that ψ be normalized then implies that $\kappa(\Delta t) = 1$, and this means effectively a specification of an “appropriate” physical environment.

The interactions, or transitions, themselves need not occur instantaneously in our interpretation. From the literature (e.g., Ref. 139) one sometimes gets the impression that in Copenhagen quantum mechanics the transitions are conceived to occur instantaneously. This is probably a remnant of the “quantum jumps” of the older (before and up to 1924) quantum theory, based mainly on the Bohr model of the atom. If the transition occurs instantaneously then the details of its temporal development lose any meaning and only the probability of its occurrence appears a reasonable question. This is what the orthodox priests of the Copenhagen interpretation assert. Schrödinger has violently opposed this view. In a discussion with Bohr in 1926 he exclaimed: “If one has to stick to this damned quantum jumping, then I regret having ever been involved in this thing”.⁽¹⁴⁰⁾ And Schrödinger kept to his view. He pointed out, for example, that “the radiating of a coherent wave train of 3 or 4 feet length, as it can be observed in an interferometer, would use up just about the average interval between two transitions”.⁽¹⁴¹⁾

4.4 The Stern-Gerlach Experiment

We will now discuss the Copenhagen postulate 2 of Sec. 4.1, namely that the measured values of any quantity are only those of the eigenvalue spectrum of the associated operator. We shall consider the case of discrete eigenvalues and shall choose the Stern-Gerlach experiment as a concrete physical example. This experiment has traditionally been regarded as the prototype of a measurement in quantum mechanics.

First let us briefly recall the facts. We consider hydrogen atoms in the ground state which move in the y direction with velocity ν through an inhomogeneous magnetic field \mathbf{B} produced by a Stern-Gerlach magnet.^(142–146) The magnet is positioned so that along the path of the atoms both \mathbf{B} and $\text{grad}B_z$ point in the z direction. This is then the “spin-reference axis”, or simply the “axis” of the apparatus. The hydrogen atom has a permanent magnetic moment $\vec{\mu}$ which comes from the magnetic dipole moment of the electron, the contribution of the proton being negligible. Therefore it is mainly the electron that interacts with the magnetic field, and it is the spin of the electron

$$\mathbf{s} = -(m/e)\vec{\mu} \quad (43)$$

that determines the precession in the magnetic field. Accordingly, in (43) m is the electron mass. We might thus just speak of electrons moving through the Stern-Gerlach magnet, and we shall occasionally do so. We note, however, that the Stern-Gerlach magnet does not work for free electrons. This is due to the Lorentz force and to the spreading of the electron wavepacket.^(147–151) If we want to perform the Stern-Gerlach experiment for free electrons we may scatter electrons by atoms.^(152,153) We call such devices Stern-Gerlach-*type* apparatuses.

Under the conditions of the performed experiments classical mechanics and electrodynamics predict that the atom when it has spent the time Δt in the Stern-Gerlach magnet will be deflected along the z direction by the angle

$$\alpha_z = p_z/p_y = \mu_z(\partial B_z/\partial z)\Delta t/p_y, \quad (44)$$

where p_y and p_z are the momentum components of the atom, and μ_z is the z component of $\vec{\mu}$. When a beam of atoms goes through the Stern-Gerlach magnet with the spins of the atomic electrons initially oriented at random μ_z can take on all values between $+\mu$ and $-\mu$, and the deflection angle can take on all values between the corresponding extreme values $\pm\mu(\partial B_z/\partial z)(\Delta t/p_y)$. On the screen behind the magnet one would therefore observe one single spot elongated along the z direction. What is actually observed, however, are two separate spots corresponding to the above two extreme values of α_z , with

$$\mu = \frac{e\hbar}{2m} = \frac{ec}{2} \lambda_C \quad (45)$$

corresponding to the electron spin value $s = \hbar/2$ in formula (43). The upper spot on the screen thus corresponds to spin-up electrons and the lower spot to spin-down electrons with respect to the axis of the apparatus.

So far our considerations of the Stern-Gerlach experiment have been independent of the interpretation since they have been formulated with beams consisting of many atomic electrons. The difficulties arise when the behavior of the individual electron wavepackets of the beam are considered. The single wavepacket develops into two coherent parts, one corresponding to spin up and the other corresponding to spin down, and it covers both the upper and the lower path in portions that can be calculated by the standard formulas. According to the orthodox view of the Copenhagen interpretation it is only in a subsequent measurement, for example, when a black spot at the proper “up” position on the screen is observed, that the packet contracts and is reduced to a pure spin-up eigenpacket. When we choose not to look at the measurement device no reduction can occur.⁽¹²³⁾

In the Copenhagen interpretation the Stern-Gerlach experiment is called a measurement of the initial z component of the spin of the atomic electron.^(154,155) In our interpretation this is different. Consider an atomic electron that is initially described by a wave function that is a product of a spatial and a spin function. Such a wavepacket never has a definite position, but it always has a definite spin component, in the sense that the wave function can always be written as a spin-up eigenfunction of the spin-component operator $s_{z'}$, with *some* axis z' , which of course need not coincide with the axis z of the Stern-Gerlach apparatus. This is connected with the fact that the group SU(2) is locally isomorphic to O(3). We may call the direction of the axis z' the spin direction of the electron before it entered the apparatus. Now, in our interpretation the Stern-Gerlach apparatus forces the electron packet to become an eigenpacket of the operator s_z , where z is the z axis of the apparatus. In other words, the Stern-Gerlach apparatus turns the spin of the electron either up or down with respect to the apparatus axis z whatever the electron’s initial spin direction z' . Because of total angular momentum conservation, the angular momentum of the apparatus is thereby also changed. This has been verified experimentally in the case of photon polarization apparatuses, which in principle function like Stern-Gerlach apparatuses.⁽¹⁵⁶⁾

In our realistic interpretation no observation is necessary to turn the electron packet into a pure spin-up or spin-down eigenpacket. Rather this occurs in some

Figure 2. Galton pin board.

physical process independently of whether or not it is observed. Where this process occurs we do not know (see Sec. 4.6). In any case we call the Stern-Gerlach *magnet* plus any additional material that might be necessary for the reduction to occur a Stern-Gerlach *apparatus*.

The Stern-Gerlach apparatus may be compared with a Galton pin board (Fig. 2) where a ball that enters it from above ends up in either box I or box II whatever its initial position. Nobody would call the observation of the ball in one of the boxes a measurement of the initial position of the ball. Accordingly, in our interpretation we do not call the operation of the Stern-Gerlach apparatus on an individual incoming electron and the observation of its respective final position on the screen a measurement of the electron's initial spin component. We may indeed use the Stern-Gerlach apparatus for such a measurement, but not without further steps: a large number of equal electrons must be given. Let the electrons enter the magnet one after the other. Then the direction of the axis of the apparatus must be varied, and from the abundance ratios of the up and down spots for the various chosen directions the original spin direction can be derived, using the standard formulas of quantum mechanics. For example, let the spin of the incoming electrons be restricted to lie in the x - z plane perpendicular to the direction of motion (y axis). The probability of inducing a spot in the “up” or “down” position, respectively, is then given by

$$P_{\text{up/down}} = \frac{1}{2}(1 \pm \cos \vartheta) \quad (46)$$

where ϑ denotes the angle between the spin axis of the incoming electron and the axis of the apparatus. The ratio of the corresponding abundances will give $(P_{\text{up}})/(P_{\text{down}}) = \tan^{-2}(\vartheta/2)$, hence ϑ and the spin direction of the incoming electrons becomes known. Alternatively, we may rotate the apparatus round the y axis until a position is obtained where only spots at the up position are observed. This signifies $\vartheta = 0$ in (46), and the apparatus axis coincides with the spin direction of the incoming electrons.

For particles with higher spin the situation is more complicated. These particles need not have a definite spin component since the expression of the general spin state as a superposition of spin-component eigenstates can no longer always be reduced to one term by a spatial rotation of the spin-reference axis. Still, the coefficients in the superposition can be determined in essentially the same way as before, and it may now be the set of these coefficients that represents what already existed before the particles entered the apparatus.

The Stern-Gerlach apparatus can also be used for a preparation of (say) spin-up atomic electrons. If the lower path is blocked off then only pure spin-up electrons leave the apparatus irrespective of their initial orientation. Only in the degenerate case where the incoming electrons all have spin down would no electron come through the apparatus. However, although we may call this operation a preparation, we are not able to predict or determine whether an individual incoming electron actually will leave the apparatus as a spin-up electron or will be absorbed inside the apparatus by the blocking material.

In our interpretation there is no superposition principle in the sense of Dirac, where it does not just mean that a linear superposition of states is again a state, i.e. that the state equations are linear and homogeneous, but rather “that whenever the system is definitely in one state we can consider it as being partly in each of two or more other states . . . in a way that cannot be conceived on classical ideas” (Dirac^(157,158)). In the situation of the Stern-Gerlach experiment this would mean that the electron even before it has entered the apparatus, when it still is in an eigenstate of $s_{z'}$ (z' = arbitrary axis), is already partly in each of the eigenstates of s_z (z = apparatus axis). We feel that this peculiar conception of Dirac’s arises from the attempt to adjust the usual view that a measurement measures what already exists to the postulate that the Stern-Gerlach apparatus operating on an individual electron is such a measurement. Indeed, if the electron were a classical gyroscope, the measurement of the place where it hits the screen behind the magnet could be considered as a measurement of its initial spin component [formula (44)]. However, the electron is not a classical gyroscope, and to describe the quantum situation in the same way as the classical situation is misleading. In our interpretation, expressing a wave function as a superposition of certain eigenfunctions in general is no more than a mathematical procedure. Only in special physical situations like the one within the Stern-Gerlach magnet, where the eigenfunctions are those of the operator that corresponds to the actual experimental arrangement, can these eigenfunctions be considered as parts of a physical wavepacket.

4.5 Criticism of von Neumann's Reduction Axiom

In this section we shall consider the last of the three Copenhagen postulates concerning the measurement as described in Sec. 4.1. This axiom reflects the most serious conceptual difficulties of the Copenhagen quantum mechanics. The realistic interpretation does not contain this axiom. However, this does not mean that there is no reduction process, it only means that this process is not restricted to occur in and only in a measurement. In this section we begin by discussing some characteristic difficulties of von Neumann's axiom and show how they are avoided in a realistic interpretation.

A typical difficulty with von Neumann's axiom is described in the example of Schrödinger's cat.⁽¹⁵⁹⁾ Consider a closed box containing a cat, a certain amount of radioactive nuclei, a Geiger counter and a cat-killing device, all protected against the cat. Circumstances are arranged so that the probability of the Geiger counter counting the decay of at least one nucleus within one hour is just $1/2$. When a decay is counted, the counter discharge triggers the cat-killing device, which consists of a hammer and a flask of prussic acid. The flask is smashed, the acid is released, and the cat is poisoned. The probability that after one hour the cat is dead is $1/2$. Since the box is closed we cannot know after an hour whether the cat is dead or alive, unless we open the box and look into it. In orthodox quantum mechanics it is allowed to describe the living cat by one wave function ψ_1 and the dead cat by another wave function ψ_2 orthogonal to ψ_1 . Then the cat in the closed box after one hour is in the state $(\psi_1 + \psi_2)/\sqrt{2}$, i.e. there is neither a dead cat nor a living cat but a superposition of both, and the superposition is only reduced to either ψ_1 or ψ_2 by our observation. Thus, in common language, it is our observation that makes the cat dead or alive, although it is not our will that determines which of the two cases actually occurs. Evidently, these features are rather hard to digest.

The cat is a macroscopic object, and in the realm of macro-objects the language of realism is spoken: the cat is either alive or dead even if we do not observe it. The radioactive nucleus is a microscopic object, and in the realm of micro-objects the language of realism is forbidden by the edict of Copenhagen. In the example there is a chain of reactions beginning in the microworld with the decay of the unstable nucleus and ending in the macroworld with the death of the cat. If both the micro- and the macro-object are described by a ψ function the character of the ψ function must change when the chain of reactions crosses the borderline between the two realms. This is another difficulty.

In our interpretation the language of realism is spoken in the microworld as well as in the macroworld, and the character of the ψ function is always that of a physical field. The cat, on the other hand, is not a wavepacket but an aggregate, and there is no reduction process of the aggregate as a whole, as we have stated in Sec. 3.3. Of course, a description of the living cat as well as of the dead cat can in principle be formulated. These descriptions could not, however, be accomplished by means of a single schematic function $\psi(\mathbf{x}, t)$, but only by extremely complicated aggregates of very many wavepackets. The essential difference in this point between our interpretation and the Copenhagen one is that if the cat dies it dies

independently of our observation, and that the dying is a process where very many wavepackets are involved, changed and possibly reduced.

Another difficulty with von Neumann's axiom is pointed out in the example of "Wigner's friend"⁽¹⁶⁰⁾: Usually the whole measurement apparatus consists of a long chain of subapparatuses (amplifier, channel analyzer, transmitter etc.) between the considered micro-object and me as the final observer. And in the most orthodox version of the Copenhagen interpretation it is my consciousness of the result that completes the measurement and effects the reduction. Now, a friend of mine may form a sub-apparatus in that chain, in that he, for example, reads off the pointer position on a display and then telephones it to me. The difficulty arises as soon as I credit the same capabilities to my friend as to me because this implies that the reduction has already taken place in the apparatus due to his consciousness. This is essentially the conflict between the Copenhagen description where the observer is the linguistic ego and any realistic description where the observer is just another physical object. The reduction is a physical process that occurs somewhere in the apparatus whether we notice it or not.

Margenau and Hill mention another strange feature of the von Neumann axiom:

Now in every normal theory of probability or random variables it is understood that a single observation cannot establish a probability distribution, a large sample of observations being required for this end. Yet the customary quantum theory, if it includes the postulate in question, claims to be a radical probability theory wherein "God plays dice" and at the same time pretends to perform the most unorthodox feat of creating, in a single act of measurement, knowledge of an entire probability distribution.⁽¹⁶¹⁾

In view of these difficulties one may ask why von Neumann's axiom has been introduced at all. What were von Neumann's reasons, for example? In his book⁽¹⁶²⁾ von Neumann refers to an experiment of Compton and Simon⁽¹⁶³⁾ in which photons were scattered by electrons at rest and where the directions of both the scattered photon and the scattered electron from the same scattering event were measured. The situation was such that, if the validity of the classical kinematical conservation laws was assumed, the azimuthal angle ϕ of the scattering plane (i.e. the angle between the plane which contains both the direction of the incoming photon and that of the momentum transfer and some reference plane which contains the direction of the incoming photon) could be determined by measuring either the azimuthal direction of the scattered electron or that of the scattered photon. These two measurements may be made in succession, and the experiment confirms that the second measurement always gives the same result for ϕ as the first one. In other words: if the system initially was in a state in which ϕ cannot be predicted with certainty then a measurement transforms it into a state in which ϕ is definitely predictable. One may accept this as an argument in favor of the reduction process; however, there is nothing in the argument that would force one to conclude that the reduction depends on the presence of an observer. The connection with an observation was perhaps conceived under the influence of some ideas of Bohr⁽¹⁶⁴⁾ quoted by von Neumann⁽¹⁶⁵⁾ (cf. also Jammer⁽¹⁶⁶⁾). If one looks at these ideas one

sees, however, that they are too vague to provide an argument against realism.

4.6 The Underlying Physical Problem

When reduction is not necessarily connected with measurement a wider range of phenomena than in Copenhagen quantum mechanics may be subsumed under the term reduction processes. Their general characteristic is that they are processes where the temporal evolution of the considered wavepacket is different from the predictions of the Schrödinger (or Dirac etc.) equation. The predicted form of the wavepacket is wider in some space (ordinary space, momentum space, angular-momentum space, etc.) than actually observed. Some narrowing process is thus missing and this is the reduction process or collapse.

Besides the Compton-Simon experiment the behavior of the atomic electron in the Stern-Gerlach apparatus is an example of a reduction process. Other examples are polarizers and beam (wavepacket) splitters of any kind, possibly together with some additional material (“registration” apparatus), and emission and scattering processes. In particular, the scattering and emission of electromagnetic radiation by atoms seem to entail some reduction processes. The usual classical and quantum mechanical calculations give more or less spherical scattered and emitted waves, whereas thermal equilibrium seems only to be possible with narrowly directed bundles of radiation (Einstein’s needle radiation⁽¹⁶⁷⁾).

Thus, the problem of the reduction process is no longer a “measurement problem”. It remains a problem, but it assumes a different form: what feature of the physical environment makes the reduction process, rather than the deterministic evolution, occur? It is now the problem of how physically (and then mathematically) to characterize the reduction process.

A physical characterization would be one in terms of current densities, field strengths, energy thresholds etc., and it would involve physical conditions and processes that are not already taken into account in the Schrödinger (or Dirac etc.) equation for a particle in an external field. Recall that reduction implies nonlinearity. In this sense quantum mechanics is incomplete, although this is a different incompleteness than that pointed out by Einstein, Podolsky and Rosen (Sec. 5). The incompleteness revealed here is a serious drawback, and the wish to avoid fully acknowledging this incompleteness presumably was another reason why the Copenhagen school ascribed those changes of the ψ function that cannot be described by a Schrödinger equation to the acts of His or Her Majesty, the Observer. In a realistic interpretation this is no longer permitted, and the problem reveals itself as a really serious one. Its final solution is not possible without changing the mathematical formalism, which is beyond the intended scope of the present work.

What our interpretation does is to reinforce the position, nowadays held openly by many physicists, that an empirical investigation of the reduction process makes sense. Actually, in recent times several experiments in this direction have been proposed or carried out.^(168–171) Generally, they are based on interference effects. Once the reduction has taken place no interference effects between the different coherent components of the original wavepacket are possible any more, simply

because there is only one component left (in the extreme case). This may serve to determine the place where the reduction occurs or is initiated. If we want to go beyond the mere localization of the reduction we may exploit the empirical finding that there exist quantum measurements where reduction does occur and that therefore the physical characterization of the reduction might be found among the physical characterizations of these measurements. A necessary condition for a measurement is certainly the possibility of an irreversible macroscopic fixation of the result, be it in the form of a developed silver grain in a photographic emulsion or in the form of a note written down by an observer. But although this is necessary for measurement, in realism it is not necessary for reduction.

4.7 Determinism and Indeterminism

It is well known that Einstein favored determinism^(172–177): “At any rate I am convinced that *he* is not playing dice”.⁽¹⁷³⁾ And this has often been considered as Einstein’s main criticism of quantum theory. However, as indicated by Einstein⁽¹⁷⁸⁾ and emphasized by Pauli in a letter to Born:

Einstein (as he explicitly repeated to me) does not consider the concept of “determinism” to be as fundamental as it is frequently held to be ... Einstein’s point of departure is “realistic” rather than “deterministic.”⁽¹⁷⁹⁾

Both determinism and indeterminism are compatible with realism. The question of determinism versus indeterminism is a matter of philosophical taste, like optimism and pessimism. We can never prove that everything is determined, but neither can we ever prove that not everything is determined, it being always possible to resort to unknown influences. Remember Freud with regard to human actions. Determinism means a program. It means the expectation that as science advances we will be able to make more and more phenomena predictable, by means of laws of nature, and that this process is infinite.

Although the question of determinism or indeterminism has nothing to do with the question of realism we nevertheless consider it here because it is given so much importance in the philosophical discussions on quantum mechanics and because the Copenhagen interpretation takes such a provocative stand. The Copenhagen interpretation takes the stand of strict indeterminism and decrees a definite limit to the described process of the deterministic program. It maintains that the probabilities in quantum mechanics are unlike those in statistical mechanics and can never be explained by some underlying determining processes that would specify the physical situation in more detail. The probability statements in quantum mechanics, according to this interpretation, are the last word. Even if we knew all the laws and all the wavepackets of the world we would not be able, in principle, to calculate the exact future result of an individual measurement. Only in some degenerate cases can we obtain probabilities that reach the value one and thus give certainty. In general, identical initial conditions do not lead to identical results.

Why does the Copenhagen interpretation assume such an extreme position? Admittedly, so far no theory has been found that could specify the hypothetical underlying processes postulated in the deterministic attitude. Thus, from the view-

point of economy of concepts it might seem reasonable to eliminate the concept of these processes altogether. This would give the additional bonus that the indeterminacy no longer points at an incapability of the quantum theorists to build a complete theory but is a property of nature.

It seems that the attitude of the Copenhagen school got additional support from von Neumann's demonstration⁽¹⁸⁰⁾ that some basic features of quantum mechanical states are incompatible with the introduction of additional hidden variables besides ψ in order to further specify the physical situations and to restore determinism. This statement seems to have been taken in the sense that no deterministic theory at all is possible. When Bell examined the case^(181,182) he found out that those basic features of the quantum mechanical states, which von Neumann postulated also for the states in a hidden-variable theory, are actually more than can reasonably be postulated in such a theory. Thus, von Neumann's proof, although mathematically correct, leaves the real question untouched and does not exclude deterministic hidden-variable theories.

On the other hand, Bell's investigations revealed that any hidden-variable theory which after averaging over the hidden variables reproduces the formulas of quantum mechanics must have a grossly nonlocal structure. A more detailed account on this specific aspect will be given in Sec. 5. Here, the essential lesson is that von Neumann's proof does not exclude deterministic theories, and, no other proof being in sight, the apodictic exclusion of determinism in the Copenhagen interpretation is unjustified. Thus Dirac wrote: "It may be that in some future development we shall be able to return to determinism, but only at the expense of giving up something else, some other prejudice which we hold to very strongly at the present time".⁽¹⁸³⁾

Let us therefore consider what a deterministic program might look like in a realistic interpretation. The probabilities and indeterminacies here are those that refer to the probabilities of the wavepacket making a transition or inducing an effect. We do not know the future formalism of quantum theory, so we do not know whether or to what extent it will be able to proceed with the deterministic program. We would, however, present some considerations that show that the deterministic program can never be fulfilled completely.

We speculate that the final form of the wavepacket in a reduction process is determined by other wavepackets in the surroundings. Usually it is assumed that they can be kept away or can be neglected so that the wavepacket is an isolated system. It is, however, hardly possible to keep away, for example, neutrinos, and it is impossible to keep away photons. A macroscopic body may easily be separated from other macroscopic bodies (put it in free space), and the influences it receives from the remaining microscopic particles of its surroundings are *relatively* very small.

On the other hand, the influences of these particles on some other microscopic particle are not always relatively small. When a photon is scattered from a billiard ball the change in the momentum of the billiard ball is negligible compared to the momentum typically considered for billiard balls, whereas when a photon is

Figure 3. The event at A can be influenced by an event at B lying outside the observer's cone of knowledge.

scattered from an electron the change in the momentum of the electron is not always negligible. Macroscopic objects are like icebergs swimming in the sea, and the microscopic particles are like the waves of this sea. Imagine the difficulties you would meet trying to influence an individual wave by handling icebergs. And the size of an iceberg does not exceed the size of a wave by as much as the size of a man exceeds the size of an atom. This last size ratio (10^{10}) is almost equal to that between the galaxy and the Sun. Seen from the microworld the macroworld of human beings and their apparatuses is of cosmic dimensions. It would be well to distinguish the difficulties that already arise from this simple fact from those that arise from the wavepacket nature of the micro-objects.

Another unavoidable influence of the surroundings comes from the special-relativistic spacetime structure.⁽¹⁸⁴⁾ Imagine a physicist at O in Fig. 3 who wants to predict exactly what will happen at A. He can do this only on the basis of what he knows, and he can at most have knowledge of the events within his past light cone. However, the event at A can obviously be influenced by events lying outside his past cone, for instance at B. Hence, as a matter of principle, future events are not completely predictable. One may restrict oneself to isolated systems and their future development. An isolated system would be one where such outside influences as that from B to A are negligible. Such a restriction may be reasonable for macroscopic bodies since appreciable influences are only to be expected from other macroscopic bodies which usually move at very low velocities and hence reach the observer's future light cone only after a very long time. For microscopic bodies, however, the isolation is much more difficult, if not impossible, to maintain.

One might try to regain predictability by extending the past light cone far backwards in time. If one assumes an appropriate cosmological model of the world the

space-time diagram will be deformed as a result of non-zero spacetime curvature and the past light cone (horizon) might eventually include all of the world, and no events such as event B in Fig. 3 will be left outside it. But even so one cannot obtain complete predictability, for one has to observe that all things consist of wavepackets. If one wants to fully describe one specific wavepacket at one particular time one can do this only with the help of macroscopic stable aggregates such as books, microfilms, disks, etc. All of these aggregates in turn consist of a large number of wavepackets. So, for noticing and registering one wavepacket many other wavepackets are needed. Only a part of the world can thus be “known”, the other part is used in representing the “knowledge”. On the other hand, some wavepacket from the registering aggregate may well influence the registered wavepacket at some later time.

Thus, the process of making more and more phenomena predictable is seen not to be infinite. Yet we do not think that we therefore should already give up the deterministic attitude. It might still be possible to extend the domain of the predictable situations at the cost of the so far unpredictable ones, or to draw a different boundary than that drawn in the Copenhagen interpretation.

5. THE EINSTEIN-PODOLSKY-ROSEN PROBLEM

5.1 The EPR Problem and Nonlocality

In this section we shall consider a problem that was first formulated by Einstein, Podolsky and Rosen (EPR) in 1935⁽¹⁸⁵⁾ and which in the course of further investigations led to the fundamental question of whether a certain nonlocality is a feature of physical reality. In the course of time the question has been made amenable to experimental test, and the results of the experiments have confirmed that nonlocal effects exist. Actually, the confirmation is not yet absolutely conclusive. The experimental results can be explained by invoking nonlocal effects, but there are still some loopholes that might permit an explanation without invoking nonlocal effects. However, our interpretation is independent of this. We conceive that there are nonlocal effects whether or not alternative explanations are also possible. If further experiments close the still existing loopholes and make the acceptance of nonlocal effects inescapable, as we think they will, people will perhaps be more willing to take an interpretation seriously in which nonlocality is an explicit basic feature. This is one reason why we discuss the EPR problem. Another reason is that it permits a concise discussion of the general consequences of nonlocal effects.

The title of the EPR paper is “Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?” The authors wanted to demonstrate that the answer is no. For a physical theory to be complete it is necessary that “every element of the physical reality must have a counterpart in the physical theory”, and reality is characterized by the following sufficient criterion: “If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity.”

EPR argue within the framework of orthodox quantum mechanics. They consider two systems, 1 and 2 (imagine two protons) that have interacted from time $t = 0$ to $t = T$, after which time there is no longer any interaction between them. Let the (exactly calculable) wave function of the combined system 1+2 after T be $\Psi(1, 2)$. The number 1 stands for all variables used to describe the first system and 2 for those of the second system. In general the function $\Psi(1, 2)$ cannot be written as a product (or a more general function) of one function $\varphi_1(1)$ of the variables 1 and one function $\varphi_2(2)$ of the variables 2, and hence we cannot describe the state in which either one of the two systems is left after the interaction. This state, according to the Copenhagen interpretation, can only be known by a subsequent measurement: Let m_1, m_2, m_3, \dots be the eigenvalues of some physical quantity (observable) M pertaining to system 1 and $u_1(1), u_2(1), u_3(1), \dots$ the corresponding orthonormal eigenfunctions. Then $\Psi(1, 2)$ can be expanded into a series of these eigenfunctions with coefficients that are functions of the variables 2

$$\Psi(1, 2) = \sum_{r=1}^{\infty} \zeta_r(2) u_r(1). \quad (47)$$

Although not necessary for the argument, for simplicity of presentation we assume that the eigenvalues are discrete. The functions $\zeta_r(2)$ are not normalized and in general are not orthogonal to each other, but this is not relevant here. Suppose that the quantity M is measured on system 1 and that the value m_7 is found. According to von Neumann's reduction axiom, after the measurement the first system is left in the state $u_7(1)$ [i.e. the first wavepacket assumes the form $u_7(1)$]. Hence the sum (47) is reduced to the single term $\zeta_7(2)u_7(1)$, and due to the simple product form of this term the second system is left in the state $\zeta_7(2)$, apart from normalization.

The set of functions $u_n(1)$ is determined by the choice of the physical quantity M . If, instead of M , we had chosen another quantity N , with eigenvalues n_1, n_2, n_3, \dots and orthonormal eigenfunctions $v_1(1), v_2(1), v_3(1), \dots$ we would have obtained another expansion

$$\Psi(1, 2) = \sum_{s=1}^{\infty} \eta_s(2) v_s(1) \quad (48)$$

where the $\eta_s(2)$ are the new coefficient functions. If now the quantity N is measured and the value n_5 is found then the system 1 is left in the state $v_5(1)$ and the system 2 in the state $\eta_5(2)$.

Therefore, as a consequence of two different measurements performed on the first system, the second system may be left in states with two essentially different wave functions. On the other hand, at the time of measurement the two systems no longer interact, that is, by definition no real change can take place in the second system as a result of anything that may be done to the first system. Thus it is

Figure 4. Arrangement for a proton spin correlation experiment

possible to assign two different types of wave functions (ζ, η) to the same physical reality, namely to system 2 after the interaction.

It is even possible to choose noncommuting operators M and N , operating in system 1, in such a way that the two sets of wave functions ζ_r, η_s of system 2 are discrete eigenfunctions of two noncommuting operators, for example of the two operators s_x and s_z of the spin component of a proton in x direction and in z direction, respectively. Such a case was first considered by Bohm.⁽¹⁸⁶⁾ Let the two protons interact at O (Fig. 4) and let the scattering proceed through an intermediate state of zero total spin (singlet spin state). The general expression (47) in this particular case becomes⁽¹⁸⁷⁾

$$\Psi(1, 2) = \frac{1}{\sqrt{2}} \left(|+\rangle^{(1)} |-\rangle^{(2)} - |-\rangle^{(1)} |+\rangle^{(2)} \right) A(1, 2) \quad (49)$$

where the spin projections up $|+\rangle$ and down $|-\rangle$ refer to an arbitrary axis. $A(1,2)$ is the spatial part and the bracket is the spin part of the wave function. Notice that the spin part follows solely from spin algebra (Clebsch-Gordon coefficients) and happens to be antisymmetric in the particle labels, independent of whether the particles are similar. In the case of similar particles the spatial part may take care of the correct symmetry.

After the interaction the protons propagate with opposite momentum $|\mathbf{p}|$ towards the observers A and B, respectively. Each observer is equipped with a Stern-Gerlach-type apparatus (e.g. a scattering device with counters; a Stern-Gerlach magnet with registering screen would not do in this case, see Sec. 4.4). The (spin-reference) axes of the apparatuses can be oriented in any direction. Observer A may thus put the axis of his apparatus either in x direction or in z direction, thereby obtaining either the x or the z spin component of the first proton. He is then in a position to predict with certainty, and without in any way disturbing the *second* proton, either the value of the x or the value of the z component of the spin of the second proton. According to the above criterion of reality both components are elements of physical reality. Therefore, the values of both must enter into the complete description of reality. On the other hand, in quantum mechanics no wave function can contain both an eigenvalue of some operator M and an eigenvalue of an operator N that does not commute with M . Therefore EPR conclude that the

quantum-mechanical description of reality by the wave function is not complete. As EPR remark, one would not arrive at this conclusion if one regarded the given criterion of reality as not sufficiently restrictive, that is,

if one insisted that two or more physical quantities can be regarded as simultaneous elements of reality *only when they can be simultaneously measured or predicted*. On this point of view, since either one or the other, but not both simultaneously, of the quantities P and Q can be predicted, they are not simultaneously real. This makes the reality of P and Q depend upon the process of measurement carried out on the first system, which does not disturb the second system in any way. No reasonable definition of reality could be expected to permit this.

Remember the burnable and soluble sugar cube in Sec. 2.5, which is another illustration of this point.

In the disputes following the publication of the EPR paper the adherents of orthodox quantum mechanics pointed out that the conclusions of EPR were only valid provided the two systems after the interaction are really independent of each other in every respect. Orthodox quantum mechanics, however, does not say this. Instead, it conceives the two particles to be inseparably incorporated into the single wave function (47), so that we cannot operate on the one particle “without in any way disturbing” the other. Only a reduction of the sum to one of its terms by means of an observation (von Neumann’s axiom) achieves a separation of the two particles.

This is independent of whether ζ, u, η and v in expressions (47) and (48) represent Schrödinger scalar wave functions or relativistic Dirac spinors or other tensors. We may consider the functions ζ, u in (47) or η, v in (48) as functions not only of the space coordinates but of time as well, with the same time variable t in all functions (cf. Sec. 3.1). Thus, the moment of the observation at the one place achieves the simultaneous reduction at the other place. Since the time of this subsequent observation is at the observer’s disposal he or she may perform it an arbitrarily long time after the interaction, so that the wave functions of the two systems can be taken to be separated from one another by an arbitrarily large distance.

On the other hand, things may be arranged so that observer A operates on system 1 such a short time before B operates on system 2 that no light signals could connect these two events. What happens at B then depends on what happened in a region that is separated from it by a spacelike distance in spacetime, hence there seems to exist some type of action-at-a-distance.

We have here a new kind of correlation between systems that have previously interacted. Before the appearance of the EPR paper this correlation had not been fully noticed. Nowadays it is also called quantum interconnectedness, quantum nonseparability, or nonlocality. Thus in taking the two systems as independent, EPR are not in accordance with quantum mechanics, hence they cannot maintain that it provides only an incomplete description of physical reality. But then the

original question “Can quantum-mechanical description of physical reality be considered complete?” is replaced by another question: “Is nonlocality a feature of physical reality?”. Einstein agreed that the EPR conclusion rests on the assumption of complete independence of the two systems after the interaction, but the assumption of non-independence or nonlocality in the form suggested by orthodox quantum mechanics appeared to him unacceptable.^(10,188)

The question is so important for our understanding of nature that, in spite of the fact that nonlocality is predicted by quantum mechanics and that quantum mechanics has been confirmed in innumerable situations, one would wish that this particular prediction be tested in special experiments. We shall consider these experiments in Sec. 5.3.

5.2 Relativity and Causality

We want to examine more closely what kind of nonlocality is predicted by the formulas of quantum mechanics. Since we use the same formulas in our interpretation, although in different situations than the Copenhagen interpretation, the discussion is relevant for our interpretation too.

In contrast to the preceding section where we argued in terms of wave functions and eigenvalues we now want to formulate the EPR action-at-a-distance or nonlocality strictly in terms of results, or probabilities of results, of experiments. For this we consider the question whether it is possible to transmit signals or messages with superluminal velocity from one person to another. Such a transmission would mean a drastic violation of relativistic causality because we may consider A’s sending the message the cause and B’s receiving it the effect, and with superluminal transmission this cause and effect could appear in reversed order of time in different Lorentz systems. We will try to construct an early-warning system.

Consider Fig. 4 of Sec. 5.1. Imagine B to be the Earth and O and A two space stations. Invaders from a distant star are expected to approach the Earth from the direction where A lies. The task of A is to inform the Earth immediately after the invaders have been seen (emergency case). For this purpose the auxiliary space station O continually emits pairs of scattered protons, say at a rate of 1 pair per second, and the protons are to pass through Stern-Gerlach-type apparatuses on station A and on the Earth, respectively. The distance between O and the Earth is made only a little larger than the distance between O and A, so that A receives its proton such a short time (which still may amount to some hours) before the Earth receives its proton that no light signal could have informed the Earth of A’s operation. During the routine cases, the apparatuses at both A and B have their axes in $+z$ direction, and in the emergency case A turns the axis of his apparatus into the $+x$ direction. One might think that this changes the probability of an up or down result in apparatus B on Earth, and from the changes in the ratio of up and down results the physicists on Earth would soon learn (before any light signal could be sent from A to B) that the invaders had been seen by A. Now, the joined probability that proton 1 in the apparatus of physicist A (in apparatus A, for short) becomes an r_A -proton (= up proton if $r_A = +1$, down proton if $r_A = -1$) and that proton 2 in apparatus B becomes an r_B -proton, according to

the formulas of quantum mechanics, is (Appendix C)

$$P(r_A, r_B | \mathbf{a}, \mathbf{b}) = \frac{1}{4}(1 - r_A r_B \cos \vartheta), \quad (50)$$

where the unit vector \mathbf{a} specifies the axis of apparatus A, \mathbf{b} that of apparatus B, and ϑ ($0 \leq \vartheta \leq \pi$) is the angle between \mathbf{a} and \mathbf{b} . Hence the probability that B observes the result r_B , whatever the result r_A , is just

$$P(r_B | \mathbf{a}, \mathbf{b}) = \sum_{r_A} P(r_A, r_B | \mathbf{a}, \mathbf{b}) = \frac{1}{2}, \quad (51)$$

independent of the axis \mathbf{a} (as well as of \mathbf{b}), and in this way the early-warning system will not work. In fact, the above arrangement cannot transmit any message, superluminal or subluminal; the superluminal case is only the most interesting aspect of this general incapability.

We may try to exploit the fact that the change of A's axis \mathbf{a} , if it does not change B's probabilities, will at least change the correlations between A's and B's results r_A and r_B , respectively. For example, when both apparatuses have their axes in $+z$ -direction, formula (50) gives

$$P(+, - | \mathbf{z}, \mathbf{z}) = P(-, + | \mathbf{z}, \mathbf{z}) = \frac{1}{2}, \quad (52)$$

that is, if A's proton is deflected in $+z$ direction ($r_A = +$) B's proton will be deflected in $-z$ direction ($r_B = -$) and vice versa, and we obtain a series of correlated results such as indicated in Table II, before the emergency case. In the emergency case, when A turns the axis of his apparatus into the $+x$ direction, formula (50) gives

Table II: Records of Results of Observers A and B Before and After the Emergency Case

$$P(+, + | \mathbf{x}, \mathbf{z}) = P(-, - | \mathbf{x}, \mathbf{z}) = P(+, - | \mathbf{x}, \mathbf{z}) = P(-, + | \mathbf{x}, \mathbf{z}) = \frac{1}{4} \quad (53)$$

so that there is no longer any correlation between the up and down results of A and of B, and the series indicated in Table II continue with uncorrelated results.

However, the physicists on Earth do not know this. They only know the results of their own apparatus, that is, the lower line of Table II, and not the upper line. Thus, they do not know the correlation of their results with those of A, still less can they realize any change in those correlations. Either of the two lines of Table II is just a random series; the probability of an up result is equal to that of a down result, before and after the emergency case. So, the early-warning system does not work this way either.

One may try more general apparatuses than just Stern-Gerlach-type ones. These also will not work. It can be shown quite generally that no faster-than-light warning system can be built with devices obeying the formulas of quantum mechanics. The proof is given in Appendix D.

On the other hand we note that the formulas of quantum mechanics only give *probabilities* for the various possible results. If A could arrange with certainty that his proton always goes into the up state (say) with respect to his axis, superluminal messages would be possible. For then, with A's and B's axes parallel, B would register only down protons. In the emergency case let A turn his apparatus upside down. From then on B would obtain only up protons, and the first of these would tell B that A has seen the invaders. For the construction of an early-warning system one might therefore try situations in the grey zone between quantum and classical physics hoping that here the probability features of quantum mechanics have already sufficiently approached classical deterministic behavior while the nonlocal features persist. An attempt has been made in that direction by Herbert⁽¹⁸⁹⁾ who used the amplification of a weak beam of light. But it was soon shown that the proposal would not work because the amplification of arbitrary states by one and the same apparatus is even at variance with the linearity of the quantum mechanical operators.^(190–193) Thus, the very theory, quantum mechanics, that predicts nonlocality also predicts that this nonlocality cannot be used for transmitting superluminal messages from one person to the other.

What, then, is the remarkable feature of formula (50), expressing nonlocality? To see this we have to consider Bell's inequality.

5.3 Bell's Inequality and Experiments

Compare the quantum-mechanical formula (50) with the formula

$$P_{\text{SC}}(r_A, r_B | \mathbf{a}, \mathbf{b}) = \frac{1}{4}(1 - \frac{1}{3}r_A r_B \cos \vartheta), \quad (54)$$

which is obtained under the “semiclassical” assumption that the two protons after the interaction are completely independent of each other and are each in a well defined spin state, with spherical symmetry on the average over many interactions (Appendix C). All our considerations so far would remain unaltered if instead of formula (50) we used formula (54). The point is that by means of Bell's inequality it can be shown that the correlations implied by formula (54) might be *locally explicable*, whereas those implied by formula (50) are locally inexplicable.^(181,194)

The Bell criterion of correlations being locally inexplicable is that the joint-probability formula leads to expectation values of the product $r_A r_B$ of the di-

chotomic variables r_A and r_B

$$E(a, b) := P(+, +|a, b) + P(-, -|a, b) - P(+, -|a, b) - P(-, +|a, b) \quad (55)$$

that violate Bell's inequality

$$K := |E(a, b) + E(a, b') + E(a', b) - E(a', b')| \leq 2 \quad (56)$$

for some choice of the parameters a, b, a', b' . (For convenience we write a and b instead of \mathbf{a} and \mathbf{b}). It is easy to see that (54) leads to $E(a, b) = -\frac{1}{3} \cos \vartheta$ and that this always satisfies Bell's inequality (56). On the other hand, formula (50) leads to $E(a, b) = -\cos \vartheta$, and this may violate (56). For example, choose vectors \mathbf{a} and \mathbf{b} that lie in planes normal to the direction of propagation of the protons, and let \mathbf{a} form the angle 0° , $\mathbf{b} = 45^\circ$, $\mathbf{a}' = 90^\circ$, and $\mathbf{b}' = 45^\circ$ relative to some standard direction. This choice results in $K = 2\sqrt{2} = 2.83 > 2$. Thus quantum mechanics implies locally inexplicable correlations and hence nonlocality.

In order to see why violation of Bell's inequality means local inexplicability we must look at the assumptions that are made in its derivation. In Appendix B it is shown that Bell's inequality is satisfied if the joint probability can be written in the form

$$P(r_A, r_B|a, b) = \int P_1(r_A|a, \lambda) P_2(r_B|b, \lambda) f(\lambda) d\lambda. \quad (57)$$

This form means that we first go to a more detailed level of description by introducing the additional variable λ into the joint probability in the integrand, and that on this level we write the joint probability in the special form⁽¹⁹⁵⁾

$$P(r_A, r_B|a, b, \lambda) = P_1(r_A|a, \lambda) P_2(r_B|b, \lambda). \quad (58)$$

The variable λ fluctuates with the probability density

$$f(\lambda) \geq 0, \quad \int f(\lambda) d\lambda = 1, \quad (59)$$

and $f(\lambda)$ and the range of λ do not depend on a, b, r_A , and r_B . Actually, λ stands for any set of variables that might be relevant. The product form (58) is more than just the separability $P(r_A, r_B|a, b, \lambda) = P_1(r_A|a, b, \lambda)P_2(r_B|a, b, \lambda)$ since in (58) the first factor does not depend on b nor the second on a . Thus, not only are the events r_A and r_B statistically independent for given λ , but also the probability that A obtains the result r_A [i.e., $\sum_{r_B} P(r_A, r_B|a, b, \lambda) = P_1(r_A|a, \lambda)$] is independent of B's parameter b ; and similarly P_2 is independent of a . There is thus no other link between A and B than the common variable λ . Of course, after the integration the probability (57) need no longer have the product form $P(r_A, r_B|a, b) = P_1(r_A|a)P_2(r_B|b)$, and correlations between r_A and r_B may arise that depend on a and b . Nevertheless, all these correlations can be explained by the underlying variable λ , and nothing forbids us to assume that the assignment

of values to λ occurs exclusively in one small spacetime region, say, in the overlap of the backward light cones of A and B. Therefore the correlations based on the probability (57) are locally explicable.

As an illustration of locally explicable correlations replace the pairs of protons in Fig. 4 of Sec. 5.1 by pairs of spaceships. Spaceship 1 of each pair will finally reach the region A of the universe. The physical conditions in this region are symbolized by the variable a , which may depend on time. When entering this region the spaceship is forced to produce either a green or a red flash. Before the start from the space port on Earth the captain of the spaceship is handed a manual with instructions on how to choose the color of the flash. The instructions may depend on the value of a at the time when the spaceship enters region A, as well as on some variable λ fixed at the start on the space port on Earth. The instructions may prescribe the flash color itself or only the special probability law to be used in selecting the color, given the values of the variables λ and a . For example, they may prescribe the use of a special roulette wheel where the relative frequency of red and black sections can be adjusted to any value prescribed by λ and a . Everything applies in the analogous manner to spaceship 2 in the independent region B with variable b . And the manuals (the parameter λ) may be different for different ships. Then these manuals may lead to (54) but they can never lead to (50).

Of course, in principle it is always possible to ascribe any spacelike correlations to causes in the past, i.e. in the common overlap of the past light cones of the events. Formula (50), for example, which cannot be brought about by means of the described manuals, can be brought about by means of other manuals in the following way. Let the manuals that are to be handed over to the spaceships at their start from the Earth prescribe not just the probabilities of the colors, but the colors themselves, as functions of the parameters a and b . Moreover, before the start of the spaceships additional manuals are sent (with light or under-light velocity) to the regions A and B with instructions on how to set the parameters a and b at the times when the ships there arrive. These manuals inform A also about the prescribed value of b , and B about the value of a , and when spaceship 1 enters region A it comes to know the values of both a and b . For every pair of spaceships different values of a and b , and of r_A and r_B , may be prescribed, so that, although everything is predetermined by the Earth, the results may appear random, and the relative frequencies of the various results r_A and r_B coincide with the quantum mechanical probabilities (50). Difficulties with this idea are encountered when we consider the original physical situation of the two protons that interact at O (Fig. 4) and subsequently enter the Stern-Gerlach-type apparatuses A and B. We must now assume that O has emitted signals that determine not only the behavior of the protons but also that of the experimenters who handle the apparatuses A and B, and this is in conflict with the free will of the experimenters.

Even if we accepted a strictly deterministic view where free will is an illusion there is an analogous difficulty. In such a view we would interpret a situation where A says that he feels his will free as a situation where his decision is determined by causes that appear random.⁽¹⁹⁶⁾ Indeed, imagine that we ask A why he had

chosen the angle of his apparatus axis to be just 7° in a particular setting. After some introspection he might give one of the following answers:

- (1) I used some kind of roulette wheel.
- (2) I saw seven cranes in the sky.
- (3) It came to my mind that my son is eagerly waiting for his seventh birthday, and so on.

Moreover, in any different setting he will usually name a different cause. This is to be contrasted with the usually quite material cause he will point out without much introspection in the case where he did not feel free: a loose screw fell between the rotating and the fixed part of the apparatus, or a colleague had installed his apparatus in the same laboratory and for some weeks this restricted the range of A's angles etc. Thus, the signal from O which is to determine A's choice must eventually show up as one of the causes from the above-mentioned variety, and yet the corresponding signal going to B invariably must inform B about A's choice. This requires a monstrous conspiracy with all the other signals around which may also contribute in influencing A and B.^(196,197) Compared with such a conspiracy a simple though spacelike link again appears the preferable assumption. Thus, if we leave such conspiracies out of consideration we may say that quantum mechanics predicts a true, locally inexplicable nonlocality.

Notice that it is irrelevant for our conclusions whether the Bell inequality can also be derived in other ways. Notice also that the question of determinism is not involved in the Bell inequality and its derivation because the parameters a, b and λ only determine the probability of an outcome, not the outcome itself. Whether this probability is reducible to some underlying constellations of additional variables is left open. Neither is the question of realism touched upon because it is left open whether the outcomes of the experiments come into existence by our observation or arise independently of the observer.

We are now prepared to return to the question, Is nonlocality a feature of physical reality? Are the formulas of quantum mechanics that predict this nonlocality confirmed in specific experiments? Many experiments have been performed by now,^(198–202) and the result is that they generally confirm quantum mechanics. Most experiments were concerned with the Bell inequality in its different but essentially equivalent forms. There are also other experiments confirming the non-local features,^(203,204) but the experiments related to the Bell inequality seem to be the most stringent ones and have been subjected to the closest scrutiny. In all of them, except in two early cases, a violation of the respective variant of the Bell inequality has been found. Moreover, the violation was exactly that predicted by the formulas of quantum mechanics. There are reasons why the two early contradicting experiments nowadays are considered less reliable. All experiments have been performed with pairs of protons that have previously interacted, or with pairs of photons from atomic cascading, two-photon decay, or parametric down-conversion, where the two independent polarizations of the photons play the role of the two spin orientations of the protons. The probability $P(r_A, r_B|a, b)$ in formula (57) or the average $E(a, b)$ are measured by means of normalized coincidence rates

and appropriate average values. The experiments were not absolutely conclusive because simplifying though very plausible assumptions had to be made. These assumptions were necessary because the filters and detectors employed were not ideal, because the two photons in the atomic-cascade experiments are not strictly antiparallel, and because of other reasons. Clauser and Horne,⁽¹⁹⁵⁾ for example, introduced the “no-enhancement assumption” which means that the photon detection probabilities [P_1 and P_2 in (58)], for every value of the variable λ , can at most be reduced but not enhanced by a polarization filter placed in front of the detector. Many of the simplifying assumptions have been gradually eliminated or reduced in their influence in the more recent experiments. For example, in all experiments before 1982 the spin-(polarization)-reference axes were fixed before the single measurements were done, hence the measurements were not separated by spacelike intervals. Therefore the result of A could, in principle, have been transmitted to B with light (or even under-light) velocity before the measurement by B had taken place and so could have influenced B’s result. Of course, in the actual experiments any mechanism that might, according to current knowledge, have permitted this was excluded; still the possibility was only excluded technically, not in principle.

In 1982 Aspect et al.⁽²⁰⁵⁾ performed an experiment in which this was excluded in principle. They used variable polarizers that jumped between two orientations in a time that was short compared with the photon transit time. In this experiment, too, a violation of Bell’s inequality and a confirmation of the quantum mechanical formulas was found. Thus, although some loopholes may still remain, on the whole the experiments provide overwhelming evidence that nonlocality is a feature of physical reality.

Finally, let us have a look at the spatial separations of the wavepackets between which EPR correlations have been observed in the experiments.

(1) In the proton-proton scattering experiment of Laméhi-Rachti and Mitig⁽¹⁵³⁾ the distance \overline{OA} in Fig. 4 was about 5 cm. The protons had a kinetic energy after scattering at O of 6 MeV, and the length of the proton packets was calculated from the lifetime of the intermediate singlet s -wave state to be 4×10^{-15} m. A proton packet of the above energy for which 4×10^{-15} m is the minimum width spreads out to an extension of 2.3 cm while its center traverses the distance of 5 cm [Appendix A, formula (A45)]. Thus, the separation between the two proton packets is about 4 times their width as measured by the standard deviation.

(2) In some experiments for testing Bell’s inequality photons from a cascade decay of excited atoms are employed. The length of the photon packets is estimated from the mean lives of the decaying levels which gives values of the order of 1.5 – 3 m. This is comparable to or even larger than the dimensions (\overline{OA}) of the apparatuses used up to 1980. In the recent experiments of Aspect et al.^(205–207) the apparatuses A and B are separated by about 13 m. This is 8 times the estimated length of the photon packets.

(3) In correlation experiments with photon pairs from e^+e^- annihilation^(208–210) the wave functions of the single photons are usually assumed to have lengths of

the order of 7 – 15 cm while the distance between O and A was up to 2.5 m. This is 16 to 35 times the packet length.

Admittedly, the lengths ascribed to the single wavepackets may be larger than assumed, in particular they may be larger than the usually adopted standard deviation Δy . The value of Δy is often calculated from Δp_y by means of the Heisenberg relation with the equality sign $\Delta y \Delta p_y = \hbar/2$. However, the equality sign can only hold for a Gaussian form of the wavepacket, and even for a Gaussian form it holds only at one instant of time; at other times the length may have spread out to values considerably larger than the minimum value. Also, the length Δy of a wavepacket is often taken to be the coherence length of the beam in which it takes part.^(153,208,209) Actually, the coherence length of the beam is of the order of a lower bound for the length of the constituent wavepackets (Appendix A).

All this, we think, is not enough to invalidate our conclusion, already mentioned in Sec. 3.1, that the observed nonlocal correlations may occur between spatially separated wavepackets. Another confirmation of this conclusion is presented in the experiment of Brendel et al.⁽¹¹⁵⁾ These authors used pairs of parametrically down-converted photons and measured the correlations in coincidence counts over distances of 55 cm. At the same time they measured the length of the single photon wavepackets and obtained values of less than 10 cm. There is thus very little overlap between the photon wavepackets. But at the same time they also obtained the high value of 87% for the visibility in the coincidence rate as a function of wavepacket separation. This cannot be explained by means of interference of wavepackets in ordinary space. And there are by now many other photon-correlation experiments that point in the same direction.^(116–118) The spatial separation between the wavepackets of a coalesced system reminds one of the separation between parts of one and the same wavepacket mentioned in Sec. 2.3.

5.4 Crucial Experiments

In this section we describe some experiments that could decide between the conceptions about systems of particles in our and in the Copenhagen interpretation. The experiments are concerned with the EPR correlations in the typical arrangement of the two protons in Fig. 4 of Sec. 5.1. In our interpretation this situation is described in the following way. When the two proton wavepackets interact they coalesce and form a coalesced system. One proton wavepacket then moves towards apparatus A and the other towards apparatus B. Then apparatus A, say, operates on its packet and modifies it. Due to the fact that the two packets form a coalesced system this modification in some Lorentz system causes a simultaneous modification of the other packet. In the last phase, A's operation leads to a splitting of the coalesced system into two completely independent single packets moving in opposite directions \overrightarrow{OA} and \overrightarrow{OB} , respectively. Thus B will meet a packet that is different from the one he would have met if the splitting had occurred due to some other interaction without A's previous operation (which is possible). The modification is such that A and B will experience the EPR correlations between the two packets as they are predicted by the formulas of quantum mechanics for similar particles. The positions of A, O and B can be arranged so that the packet 2

at the time of A's operation is already close to B and the interval of time between A's operation and the arrival of packet 2 at B is so short that no light signal could have travelled in that time from A to B. In this case the correlations occur between spacelike events.

Now, since these correlations are a consequence of coalescence and since coalescence occurs only between similar wavepackets we predict that no EPR correlations will occur between dissimilar particles. In fact, EPR correlations so far have only been observed between similar particles, namely with pairs of photons and with pairs of protons. In these cases we predict the same results as the Copenhagen interpretation, except for the following restriction.

The strange "telepathy" or "quantum nonseparability" exhibited in the EPR situations with similar particles is weaker in a sense in our interpretation than in the Copenhagen one. In Copenhagen quantum mechanics, the two particles are interrelated until a reduction occurs. The reduction occurs only in an observation. The observation is at the observer's disposal, and he may perform it an arbitrarily long time after the interaction, when the particles are separated from each other by an arbitrary long distance. In our interpretation we do not speak of EPR nonlocality or telepathy between different objects but only within one and the same object, namely the coalesced system. The range of the nonlocality is limited by the extension of the coalesced system, and this, in realism, is an objective physical criterion. Moreover, there are the objective processes of splitting and reduction, induced by other wavepackets in the surroundings, and due to these processes the chance of the coalesced system to spread out over a large region decreases with its size.

These considerations lead us to the following proposals.

(1) The first is a variant of an experiment that has already been performed, namely the proton-proton scattering experiment.⁽¹⁵³⁾ In that experiment protons of 13 MeV kinetic energy were scattered by a hydrogen target. After scattering each of the two protons was slowed down to 6 MeV and then entered a spin analyser in which it was scattered a second time by a carbon foil and then registered by one of two detectors. These detectors lay in a plane that contained the direction of the proton that entered the analyzer, each detector forming a fixed angle of 50° with that direction. In each of the two analyzers the plane with the detectors could be rotated around the incoming-proton direction. The coincidences of the one analyzer with the detectors of the other (N_{LL}, N_{LR} etc.) were counted, and the expression

$$(N_{LL} + N_{RR} - N_{RL} - N_{LR}) / (N_{LL} + N_{RR} + N_{LR} + N_{RL}),$$

after some corrections, is $E(a, b)$ of formula (55). This was compared with the maximum possible value compatible with the Bell inequality. Thus a violation of Bell's limit, in accordance with the quantum mechanical formula $E(\mathbf{a}, \mathbf{b}) = -\mathbf{a} \cdot \mathbf{b} = -\cos \vartheta$ could be demonstrated. If this experiment were to be repeated with the incoming proton replaced by a neutron, say, forming a singlet state with the proton, we predict that in place of the EPR correlations implied by formula (50) weaker

correlations would be observed. The Copenhagen interpretation prescribes the use of formula (50) in both cases.

We conjecture that the weaker correlations will be given by formula (54), which follows from the assumption that after each interaction either of the two particles is in a well defined spin state (Appendix C). This is plausible here because any spin- $\frac{1}{2}$ particle that arises from a spin-singlet state (but not a particle with higher spin) is always in an eigenstate of the spin component operator with respect to *some* axis, and the particular axis does not matter. Physically, using formula (54) is equivalent to assuming that the reduction process occurs already during or immediately after the interaction. In the Copenhagen interpretation formula (54) would apply only if a measurement had taken place immediately after the interaction since here only measurement entails reduction. Notice that both formulas, (50) and (54), are met in either interpretation. The difference between the interpretations is the prescription of which formula is to be used in which physical situation.

Even if the weaker correlations are not known exactly this does not necessarily rule out their experimental distinction from the corresponding EPR correlations. It is always possible to choose such properties (observables) that the EPR correlations between them are separated from the non-EPR correlations by a finite gap in the Bell inequality (56), which may be as large as 41%.^(211,212)

However, it must be noted that in the special case where the incoming proton in the proton-proton scattering experiment is replaced by an antiproton EPR correlations might still arise. Protons and antiprotons are dissimilar and no coalescence should occur. Nevertheless, in the mathematical treatment of electron-positron scattering in quantum electrodynamics virtual-annihilation terms appear, and these terms have the same mathematical structure as exchange terms.^(213,214) As exchange terms arise from symmetrization and hence from coalescence, the same must be expected of the annihilation terms. And we expect that this carries over to any particle-antiparticle pair and that in any such pair coalescence and hence EPR correlations may occur.

(2) Second, we suggest to consider the spin correlations between the proton and the neutron arising from the breakup of a deuteron. The breakup may be induced by photons with (laboratory) energies E_γ above the threshold of 2.226 MeV. The total cross section for this photodisintegration process has a maximum of 2.5 mb at about 4.4 MeV. Up to energies of about 2.4 MeV formation of the spin singlet state (M1 transition, $^3S \rightarrow ^1S$) prevails, then that of the triplet state (E1 transition, $^3S \rightarrow ^3P$).^(215–218) The singlet state can be treated by means of the simple formula (50) or $E(\mathbf{a}, \mathbf{b}) = -\mathbf{a}\mathbf{b}$. In the triplet states we have⁽²¹⁹⁾ $E(\mathbf{a}, \mathbf{b}) = a_z b_z$ in $|1, +1\rangle$ and $|1, -1\rangle$, and $E(\mathbf{a}, \mathbf{b}) = \mathbf{a}\mathbf{b} - 2a_z b_z$ in $|1, 0\rangle$ where a_z and b_z are the components of \mathbf{a} and \mathbf{b} in some preferred direction. This direction, that is, the exact wave function or density matrix of the triplet state has also to be known, so the triplet case is more complicated than the singlet case. One may therefore restrict oneself to the singlet case, that is, to E_γ near threshold and to forward and backward directions with respect to the direction of the gamma rays. A small admixture of up to about 10% of the triplet case appears acceptable

because it still gives quantum mechanical values of $|E(\mathbf{a}, \mathbf{b})|$ that can be larger than the maximum possible value compatible with the Bell inequality, and one may proceed as in the proton-scattering experiment. The main difficulty with this experiment then, as far as I can see, is the comparatively low flux of presently available gamma ray sources and the low overall polarimeter efficiencies, which give too few coincidences in a reasonable time. Perhaps it will be possible to find more efficient polarimeters for nucleons whose energy is lower than in the proton-scattering experiment, employing gamma ray energies closer to the threshold.

In the spin-singlet case Copenhagen quantum mechanics prescribes the application of formula (50), and this can lead to a violation of Bell's inequality. In our interpretation formula (50) must not be applied, and our conjecture is formula (54), which always satisfies Bell's inequality.

(3) Even for systems of similar particles the Copenhagen and the realistic interpretation differ in some experimental predictions. For example, the coalesced system $\Psi_{SA}(\mathbf{x}_1, \mathbf{x}_2, t)$ when it expands from the place O of interaction towards the observers A and B may interact with other wavepackets around. There may thus occur some unobserved spurious scattering events with the effect that Ψ_{SA} is modified before it reaches A or B so that the correlations between the results at A and at B are decreased. Now, in the realistic interpretation these scattering events may have the additional effect of splitting the coalesced system Ψ_{SA} , and this leads to an additional decrease in the EPR correlations. In the Copenhagen interpretation this splitting does not occur since here the splitting is connected with a reduction and the reduction occurs only in the measurement in apparatus A or B. The decrease in the EPR correlations is thus stronger in the realistic than in the Copenhagen interpretation.

One could study this decrease by putting a variable amount of material between O and A (or B) or by varying the distance \overline{OA} , in the case that there is vacuum (i.e. a rarefied gas) between O and A. Unlike the experiments in (1) and (2) this type of experiment does not allow a quantitative prediction since the details of the splitting interactions are unknown at present. Reversely, one could use observed deviations from the predictions of the Copenhagen quantum mechanics in this situation to specify the splitting interactions, in the sense of the program mentioned at the end of Sec. 4.6 to define the reduction process in physical terms. Experiments with variation of the distance \overline{OA} have been performed with photon pairs from e^+e^- annihilation.⁽²⁰⁸⁻²¹⁰⁾ In Ref. 208 a change in the correlations was reported, whereas the later experiments^(209,210) did not detect any significant change. One might, however, suspect that the considered distances of less than 2.5 m were not large enough. Experiments with a variable amount of material interposed between O and A have also been proposed by Jauch.⁽²²⁰⁾ To my knowledge they have not been performed so far.

6. QUANTUM STATISTICS WITH WAVEPACKETS

6.1 Field Quantization

As we have already pointed out in Sec. 2.4 on normalization and quantization the Schrödinger equation in ordinary $(3+1)$ -dimensional space is a classical field equation, and the discrete eigenfunctions following from imposing the usual normalizability and uniqueness conditions are no more than the standing waves of classical physics. This is just a consequence of de Broglie's idea of matter waves, and beyond this no new element is introduced. We would thus not call that appearance of discrete eigenvalues and eigenfunctions real quantum effects but would reserve this denomination to effects that cannot be explained in the described way. At first, Schrödinger seems to have believed that there are no such other effects, but he was opposed by Heisenberg^(221,222) who pointed, among other things, to the photoelectric effect and to the Planck radiation law. Regarding the photoelectric effect, we have shown in Sec. 2.3 that indeed something new beyond the classical wave description is needed in order to account for the missing delay in photon counts, and this is the internal structurelessness. Regarding the Planck law we shall see that Schrödinger was essentially right.

Schrödinger's equation in its general form is not an equation in ordinary space but in $(3N+1)$ -dimensional configuration space, and this goes beyond de Broglie's conception. The Schrödinger or de Broglie function $\psi(\mathbf{x}, t)$ in ordinary space in itself does not tell us whether it refers to one or to more particles or to particles at all. It is just a field and expresses only the wave aspect. With the introduction of the configuration-space function $\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N, t)$, however, the number N of particles or quanta is explicitly introduced. This means real quantization, and Ψ expresses both the wave and the particle aspect.

In 1927 in his fundamental paper on the quantum theory of the emission and absorption of radiation Dirac⁽²²³⁾ derived, among other things, the Einstein A and B coefficients and hence Planck's law by means of a new procedure, which had first been introduced by Born and Jordan⁽²²⁴⁾ and in the famous three-man work on matrix mechanics by Born, Heisenberg and Jordan.⁽²²⁵⁾ The procedure consisted in turning some canonically conjugate variables of the Hamilton formalism into operators satisfying the canonical commutation relation, such as between position and momentum. Here the now familiar creation and annihilation operators a^\dagger and a (in present notation) showed up for the first time. In the same paper Dirac showed that this procedure was equivalent with Schrödinger's configuration space treatment with symmetrical wave functions. The equivalence was subsequently elaborated and extended by Jordan, Klein and Wigner⁽²²⁶⁻²³⁰⁾ to include antisymmetric wave functions, i.e. fermions, in which case anticommutators were to replace the commutators. In 1932 Fock⁽¹⁰⁸⁾ gave a lucid summary of these developments, and he showed that the restriction to a configuration space of fixed dimension can easily be overcome. This he made particularly clear by casting Schrödinger's configuration-space formalism into the form of the Fock, occupation-number, or N representation,^(108,110) which we have already discussed in Sec. 3.1. In this representation the total number of particles no longer appears explicitly,

and this makes it possible to apply the formalism to systems in which the total number of particles is not conserved, as for example to the photons in a cavity.

The method of commutation relations was then further developed into a comprehensive scheme by Heisenberg and Pauli.^(231,232) They no longer derived the commutation relations from configuration space but set out from ordinary-space fields $\psi(\mathbf{x}, t)$ and introduced the commutation relations by way of postulate. In this way the quanta of the fields (i.e. the particles) arise from interpreting the operator $a^\dagger a$, which has only non-negative integer eigenvalues, as a particle-number operator. Moreover, and most important, they extended the formalism to include Lorentz invariant interactions and hence retardation between the similar particles. Retardation effects cannot be taken into account in Schrödinger's configuration space, so the two schemes are no longer equivalent. We call Heisenberg's and Pauli's scheme canonical quantization. In it the well known difficulties with the diverging integrals, irreparable by simple normal ordering, began. Thus, in my opinion, this is where something went wrong with the relativistic formulation of quantum theory, and I suspect that this is related to the general negative attitude towards nonlocality at that time, as reflected, for example, by Pauli's classification⁽⁸⁷⁾ of Landau's and Peierls' nonlocal density [formula (12)] as "unnatural". It appears that the problem is how to build into the theory the right dose of nonlocality at the right place, as we have already mentioned at the end of Sec. 3.1.

Thus here we stop, and we conclude our treatise with the re-interpretation of the nonrelativistic field-quantization formalism as far as it is equivalent with Schrödinger's configuration-space formalism, including symmetrization and particle non-conservation. Heisenberg's and Pauli's scheme and the subsequent relativistic developments are no longer dealt with. Our treatment still includes quantization of the electromagnetic radiation field, that is photons. There are no retardation effects between the photons, because there are no interactions between them.

In the following sections we shall derive the Bose, Planck and Fermi distributions as well as the corresponding fluctuations, by speaking of alteration, condensation and decondensation of wavepackets. The configuration-space wave functions which are behind all this are symmetrized and thus describe a coalesced system of similar wavepackets. As we have discussed in Sec. 3.1 this may lead to EPR correlations between the wavepackets. These correlations are, however, not relevant in the questions discussed here, and we may treat the packets as independent. Nevertheless, *some* correlations are already implicit in our considerations, namely those that make themselves felt as condensation and Pauli exclusion.

In general the one-particle basis functions $\varphi_{r_i}(\mathbf{x}, t)$ in the expansion (19) of the function Ψ_{SA} are merely a mathematical system of functions and cannot be identified with independent physical wavepackets. This does not exclude that in special physical situations they can so be identified and also that they can exist independently, without being part of a coalesced system. Thus, in the following sections where the $\varphi_{r_i}(\mathbf{x}, t)$ are energy eigenfunctions or narrow superpositions of these they may effectively be considered as independent wavepackets, much as the up and down parts of the wavepacket describing the spin- $\frac{1}{2}$ particle in the

Stern-Gerlach magnet may be considered as independent packets as far as their interactions with the magnetic field are concerned.

6.2 The Many Aspects of the Condensed Wavepackets

We consider similar particles of mass m in a cavity of volume V at temperature T , and we write the general quantum statistical distribution function in the well-known form

$$N(p, T)dp = \frac{4\pi V p^2 dp}{h^3} \times \left\{ \exp \left[\left(\sqrt{p^2 c^2 + m^2 c^4} - \mu \right) \frac{1}{kT} \right] \pm 1 \right\}^{-1} \quad (60)$$

where $N(p, T)dp$ means, in the usual interpretation, the time averaged number of particles in V whose absolute value of momentum lies in the interval dp about p . The plus sign refers to fermions and the minus sign to bosons. μ is the chemical potential [fugacity $z = \exp(\mu/kT)$]. In the special case of photons we have $m = 0$, $p = h\nu/c$, $\mu = 0$, and formula (60) with the minus sign reduces to the Planck distribution for polarized radiation.

We have written the distribution (60) as the product of two factors. The first factor is

$$g_p = \frac{4\pi V}{h^3} p^2 dp = \frac{4\pi V \epsilon \sqrt{\epsilon^2 - (mc^2)^2}}{h^3 c^3} d\epsilon \quad (61)$$

where $\epsilon = \sqrt{p^2 c^2 + (mc^2)^2}$ is the total energy of a particle. In this section we will consider only this factor; the second factor will be considered in the next section. In the special case of photons (61) becomes

$$g_\nu = 4\pi V \nu^2 d\nu / c^3, \quad (62)$$

and in this case it has a long history:

In 1899 it was calculated by Planck⁽²³³⁾ as the proportionality factor between the mean energy of electromagnetic radiation in V and $d\nu$ and the mean energy of a charged oscillator with radiation damping. In 1900 and 1905 Rayleigh⁽²³⁴⁾ and Jeans⁽²³⁵⁾ considered the factor as the number of degrees of freedom of the ether inside the cavity, this number in turn being considered equal to the easily calculable number of eigenvibrations (modes of vibration) within $d\nu$ of the ether. In 1914 von Laue⁽²³⁶⁾ decomposed the cavity radiation into mutually independent radiation bundles, each converging to its focal region and then diverging. To these bundles he attributed degrees of freedom and obtained (62) as the sum of the degrees of all these bundles (see below). Bose, in his famous paper of 1924,⁽²³⁷⁾ considered the factor (62) as the number of phase-space cells of size h^3 . Such cells had already been considered by Planck in 1906⁽²³⁸⁾ in the special case of harmonic oscillators. In 1925 Landé⁽²³⁹⁾ called those of von Laue's radiation bundles that had just one degree of freedom elementary light-quantum bundles or just quantum bundles, and he proposed to identify these with Bose's quantum phase-space cells.

In present-day quantum mechanics (61) is the number of eigenvalues of the Hamilton operator for a free particle in V that fall into the energy interval $d\epsilon$ that

corresponds to dp . Each eigenvalue is multiply counted according to its order of degeneracy. In other words, (61) is the number of eigenstates in V and dp . For photons we are thus effectively back at Rayleigh's and Jeans' determination. In quantum mechanics (61) holds, however, for any kind of particle, not just photons, because de Broglie waves are associated with each kind of particle. Finally, in canonically quantized radiation theory (62) is the number of oscillators. But in contrast to Planck's oscillators, which represent atoms interacting with the radiation field, these oscillators are to represent the field itself, a point of view that had already been indicated by Ehrenfest in 1906.⁽²⁴⁰⁾

Now we want to show that (61) or (62) can also be taken as the number of (condensed) wavepackets in the cavity covering the momentum interval Δp . For this purpose we employ the fact that (62) is the total number of degrees of freedom of von Laue's radiation bundles and that a bundle of F degrees of freedom may be taken to consist of F wavepackets. Von Laue defines the number F of degrees of freedom of a bundle of length l (from wall to wall of the cavity), spectral range $d\nu$ (equal to the spectral range of the radiation considered), convergence half angle α and focal cross section A with the help of the theory of optical resolving power and the counting of Fourier coefficients. He arrives at the expression

$$F = \frac{Ald\nu}{ac} \quad (63)$$

where a is the minimum focal area that is possible for a bundle of convergence angle α [cf. formula (A39) of Appendix A with $\alpha = v_{sx\infty}/c$]. Now imagine that the bundle of F degrees of freedom consists of a stream of wavepackets, all moving parallel to the axis of the bundle and going side by side through its focal area. The convergence of the bundle to the focal plane and its subsequent divergence comes about by the contraction and subsequent spreading in the transverse direction of each of these wavepackets, assuming that all packets have their minimum transverse extension in the focal plane. In front of and behind the focal plane the wavepackets may overlap in the lateral direction. We then write von Laue's degrees of freedom F as the product of three factors: $F = N_1 N_2 N_3$ where $N_1 = d\nu/\Delta\nu$, $N_2 = A/a$, and $N_3 = l/(2\Delta y)$. Each factor is the ratio of some quantity relating to the bundle divided by the corresponding quantity relating to the packets. $\Delta\nu$ is the frequency range of a wavepacket of total length $2\Delta y$. $\Delta\nu$ is related to Δy by $\Delta\nu = rc/(4\pi\Delta y)$, ($r \geq 1$), which follows from the Fourier reciprocity (Heisenberg) relation $\Delta y \Delta p_y = r\hbar/2$ with $\Delta p_y = (h/c)\Delta\nu$. With this relation we obtain $N_1 N_2 N_3 \leq (2\pi/r)F$. The number r may be set equal to 2π because the packets in the cavity do not all have their minimum phase-space extension ($r = 1$). Moreover, there is always some degree of arbitrariness in the exact definition of the widths Δy , Δp_y etc., resulting in some arbitrariness in r . Thus we take $r = 2\pi$ to mean the average extension of the wavepackets in the cavity with an appropriate definition of the widths, and with this we obtain

$$N_1 N_2 N_3 = F. \quad (64)$$

Now, by the definitions given above N_1 is the number of spectral types of wavepackets in the bundle, as defined by their individual frequency ranges $\Delta\nu$ (colors), N_2 is the number of wavepackets of a particular spectral type that go side by side and N_3 those that go one after the other through the focal area of the bundle. It follows that the product $N_1 N_2 N_3$ is equal to the total number of wavepackets that make up the bundle, and the relation (64) means that this number is equal to the number of degrees of freedom of the bundle. Thus the total number of wavepackets can be identified with the total number (62) of degrees of freedom, or of von Laue's bundles if we imagine the radiation to consist only of elementary bundles. This not only holds for photon packets but for packets of any kind. The considerations by von Laue can easily be extended to matter waves. One just has to replace $d\nu$ by $(c/h)dp$ in F and N_1 , and the same result obtains.

Now, the discrete energy values $n h \nu$ may be attributed to each of Planck's oscillators or Jeans' degrees of freedom,⁽²⁴¹⁾ and n quanta may occupy each of Bose's cells. These are then our *condensed* wavepackets representing n quanta. Empty wavepackets, without a quantum, are also included in formula (61). This is a convenient means of indicating that there is space left for more wavepackets to show up in V and dp (cf. Bose⁽²³⁷⁾ and Schrödinger⁽²⁴²⁾) Equivalently, one may say that (61) is the maximal possible number of existing (non-empty) wavepackets.

The condensed wavepackets also resemble the degenerate light pulses of Mandel.⁽²⁴³⁾ Mandel introduced the degeneracy parameter δ , meaning the average number of photons in a light beam that are to be found in the same cell of phase space. He expressed the phase space volume with the help of a certain coherence volume defined in the theory of optical coherence. We obtain the same formulas when we take the phase-space volume to be the product of the ranges of the wavepacket at the moment of its minimum extension (Appendix A).

Finally we want to point out to the "light molecules", "quantum multiples", " n -quantum rays" and "radiation bundles" first mentioned parenthetically by Joffé (1911)⁽²⁴⁴⁾ and considered more closely by Ishiwara (1912),⁽²⁴⁵⁾ Wolfke (1921),⁽²⁴⁶⁾ de Broglie (1922),⁽²⁴⁷⁾ Bothe (1923, 1924)⁽²⁴⁸⁾ and especially by Schrödinger (1924)⁽²⁴²⁾ and Bothe (1927).⁽²⁴⁹⁾ These authors noticed that the Planck distribution can be written in the form of a sum and that the n -th term can be interpreted as a contribution from objects that are composed of n light quanta (see formula (76) below). It seems that these ideas retreated under the blow of Dirac's quantization of radiation in 1927,⁽²²³⁾ but it is seen that they also strongly resemble our condensed wavepackets.

The wavepackets, not the single quanta (if we were to take these for a moment as entities of their own) are the statistically independent objects, and condensation of wavepackets is our means of expressing the "mutual influence of the molecules [i.e. quanta] which for the time being is of a quite mysterious nature" mentioned by Einstein in 1925.⁽²⁵⁰⁾

6.3 The Balance Relation

Now we turn to the second factor in (60)

$$\left\{ \exp \left[\left(\sqrt{p^2 c^2 + m^2 c^4} - \mu \right) \frac{1}{kT} \right] \pm 1 \right\}^{-1}. \quad (65)$$

We want to derive this factor by means of Einstein's method of balance relations between transition rates,⁽¹⁶⁷⁾ although with some significant modifications. These modifications are necessary because Einstein's original treatment of 1917 did not explicitly take into account that photons are bosons and not fermions; it would give the same distribution function in both cases. Of course, the Fermi distribution was published only in 1926.⁽²⁵¹⁾ The fact that photons are bosons can only be taken into account when phase-space regions or energy intervals are subdivided into those fundamental units that are counted by formula (61). It is not enough to consider the number of photons in a given energy interval, as in Einstein's procedure of 1917, but one must further specify how the photons are distributed over the various fundamental units within this interval. And these same fundamental units, in the case of fermions, must not be occupied by more than one quantum, due to the Pauli principle. The fact that Einstein did obtain the Planck distribution, in spite of not accounting for the subdivision into fundamental units, is due to the special way he formulated the balancing equations. We shall return to this point below.

With the subdivision into fundamental units we shall be able to derive both the Bose and the Fermi distribution on an equal footing by means of the method of balancing equations. The Fermi distribution has already been obtained in the framework of this method by several authors, though in a different way.^(109,252–257) Our procedure is, in fact, inspired by the comprehensive treatment by Oster⁽²⁵⁴⁾ and the remarkable paper by Bothe.⁽²⁴⁹⁾ The general mechanism in all these works is exchange of quanta between fundamental units. These units have sometimes been taken to be the discrete energy states of atoms or oscillators which emit and absorb photons. We emphasize, however, that the fundamental units are not restricted to discrete energy states. That they may well be small but finite energy intervals, centered about any energy values, had already been pointed out by Pauli⁽²⁵⁸⁾ and Einstein and Ehrenfest⁽²⁵⁹⁾ when considering photons scattered by free electrons in the Compton effect.

In our interpretation the fundamental units are the wavepackets and we are going to consider processes taking place between these. As we have discussed in Sec. 3.1 this is our interpretation of the change in "occupation numbers", which in the occupation-number representation is described by means of the creation and annihilation operators. The one-particle basis functions in the expansion (19) now are energy eigenfunctions or narrow superpositions of these. Specifically, we consider two types of elementary processes. Either of them is decomposed in alteration, condensation and decondensation. Alteration changes the energy of a wavepacket, whereas condensation and decondensation change the number of quanta it represents. Thus the first type is:

Figure 5. Scheme of the considered processes (1) and (2) between the states $|a\rangle$ and $|b\rangle$

(1) An s -quantum wavepacket of kind 1 which represents s quanta in the energy interval $\epsilon_1^i \dots \epsilon_1^i + d\epsilon_1^i$ (an s -packet in $d\epsilon_1^i$, for short) decondenses into an $(s - n)$ -packet and an n -packet in $d\epsilon_1^i$. The n -packet exchanges energy and momentum with an n' -packet of kind 2 in $d\epsilon_2^i$ whereby it is altered and goes into the energy interval $d\epsilon_1^f$ and then condenses with an r -packet in $d\epsilon_1^f$ to form an $(r + n)$ -packet in $d\epsilon_1^f$. Simultaneously an s' -packet of kind 2 in $d\epsilon_2^i$ decondenses into an $(s' - n')$ -packet plus an n' -packet in $d\epsilon_2^i$. The n' -packet is altered in an interaction with the n -packet in $d\epsilon_1^i$ whereby it goes into the interval $d\epsilon_2^f$ and then condenses with an r' -packet in $d\epsilon_2^f$ to form an $(r' + n')$ -packet in $d\epsilon_2^f$. A graphical scheme is presented in Fig. 5. Conservation of energy requires

$$n (\epsilon_1^i - \epsilon_1^f) = n' (\epsilon_2^f - \epsilon_2^i). \quad (66)$$

The energy intervals $d\epsilon$ are chosen so that they correspond to the respective intervals dp in absolute value of momentum used in formula (60), i.e. $d\epsilon = (d\epsilon/dp)dp$. Effectively, if we may use here the picture of quanta as standard portions of water, an s -packet in $d\epsilon_1^i$ gives n quanta to an r -packet in $d\epsilon_1^f$, and an s' -packet in $d\epsilon_2^i$ gives n' quanta to an r' -packet in $d\epsilon_2^f$. Thus, process (1) leads from the state $|a\rangle$ characterized by the 4 packets which represent s, r, s' and r' quanta respectively, to some state $|b\rangle$ characterized by the 4 packets that represent $s - n, r + n, s' - n'$ and $r' + n'$ quanta respectively. The probability of such a transition is denoted by W_1 . The process comprises most particular physical situations that lead to Fermi or Bose distributions as special cases.

There is no interaction between photons, so here we need the second kind of wavepacket (atoms, electrons etc.). Electrons interact with each other, and kind 1 and kind 2 may be the same. In the Boltzmann case there is no change in the number of quanta but only an alteration in the energies of the wavepackets. This can be described by putting $r = r' = 0, n = s, n' = s'$ in in the scheme of Fig. 5 and dropping $p(0, \epsilon)$ and $q(0, \epsilon)$ from the Eqs. (67) and (68) below, which then lead to the Boltzmann distribution for the wavepackets (of energy ϵs).

(2) The second type of processes is this: an $(r+n)$ -packet of kind 1 in $d\epsilon_1^f$ (which need not be the same packet as that at the end of process (1)) decondenses into an r -packet and an n -packet. The n -packet exchanges energy and momentum with an n' -packet of kind 2 whereby it goes into the energy interval $d\epsilon_1^i$ and then condenses with an $(s-n)$ -packet of that interval. Simultaneously an $(r' + n')$ -packet of kind 2 in $d\epsilon_2^f$ decondenses into an r' -packet and an n' -packet. The n' -packet is altered in an interaction with the n -packet of kind 1 whereby it goes into $d\epsilon_2^i$ and then condenses with an $(s' - n')$ -packet. Conservation of energy is again guaranteed by Eq. (66).

The initial (final) s, r, s' and r' -packets of process (2) have the same momenta etc. as the final (initial) s, r, s' and r' -packets of process (1) and differ from those only by spatial translations. Thus, process (2) goes back from state $|b\rangle$ to state $|a\rangle$. The probability of this transition is denoted by W_2 .

Process (2) is not the time reversed (“converse”) process to process (1), but may be called the “reverse” process, after Dirac.⁽²⁶⁰⁾ Only the reverse process can lead to statistical equilibrium,⁽²⁵⁹⁾ but only the converse process is suggested to exist and to occur at the same rate as the original process on account of the general principle of time-reversal invariance of basic processes. Now, in an isotropic medium the reverse process can be obtained from the converse one by successive reflections in three mutually perpendicular mirrors at rest relative to the system as a whole, and their frequencies of occurrence must be equal.⁽²⁶⁰⁾ Thus, in statistical equilibrium the two processes (1) and (2) also occur at equal rates. We now consider these rates. Rate 1 is the mean number of processes (1) that occur per second in the volume V . According to the above-given description it should be equal to

$$p(s, \epsilon_1^i) d\epsilon_1^i p(r, \epsilon_1^f) d\epsilon_1^f q(s', \epsilon_2^i) d\epsilon_2^i q(r', \epsilon_2^f) d\epsilon_2^f W_1 \quad (67)$$

where $p(s, \epsilon_1^i) d\epsilon_1^i$ is the mean (time averaged) number of s -packets of kind 1 in V that represent quanta in $d\epsilon_1^i$, $q(s', \epsilon_2^i) d\epsilon_2^i$ is the mean number of s' -packets of kind

2 in V and $d\epsilon_2^i$, and so on. Analogously, for process (2) the rate is

$$p(s - n, \epsilon_1^i) d\epsilon_1^i p(r + n, \epsilon_1^f) d\epsilon_1^f q(s' - n', \epsilon_2^i) d\epsilon_2^i q(r' + n', \epsilon_2^f) d\epsilon_2^f W_2, \quad (68)$$

and the two rates (67) and (68) have to be equated.

Now, according to the preceding section our wavepackets and hence the states $|a\rangle$ and $|b\rangle$ mean pure states of quantum mechanics, and since the probability of a transition in quantum mechanics is the same for a process that goes from $|a\rangle$ to $|b\rangle$ as for a process that goes from $|b\rangle$ to $|a\rangle$ (Hermitean operators), the probabilities W_1 and W_2 are equal and disappear from the balancing equation. We shall thus obtain the statistical distribution functions *without using any special property of the transition probabilities*. The differentials $d\epsilon$ also cancel, and our balance relation acquires the simple and symmetric form

$$p(s, \epsilon_1^i) p(r, \epsilon_1^f) q(s', \epsilon_2^i) q(r', \epsilon_2^f) = p(s - n, \epsilon_1^i) p(r + n, \epsilon_1^f) q(s' - n', \epsilon_2^i) q(r' + n', \epsilon_2^f). \quad (69)$$

The relation is reminiscent of the relation for chemical equilibrium between several kinds of molecules. It still comprises the Bose and Fermi cases. There is no spontaneous emission term, i.e. one that would be independent of the number of wavepackets. A general solution is

$$p(s, \epsilon) = a(\epsilon) \exp[-(b\epsilon - c)s] \quad (70)$$

$$q(s, \epsilon) = a'(\epsilon) \exp[-(b\epsilon - c')s].$$

[Insert and use (66)]. Notice that only the parameter b (which shortly will be identified with $1/kT$) is the same for the two kinds of packets. In any other respect the distribution function for kind-1 packets is independent of the distribution function for kind-2 packets.

The parameters a and b are obtained via the thermodynamic relations $\partial S / \partial E_{\text{tot}} = 1/T$ and $\partial S / \partial N_{\text{tot}} = -\mu/T$ where

$$S = k \ln \prod_{\{d\epsilon_i\}} \frac{g_p!}{[p(0, \epsilon_i) d\epsilon_i]! [p(1, \epsilon_i) d\epsilon_i]! \dots} \quad (71)$$

is Bose's or Natanson's⁽²⁶¹⁾ formula (in our notation) for the entropy of the total system, with $N_{\text{tot}} = \sum_{\{d\epsilon_i\}} N dp$ and $E_{\text{tot}} = \sum_{\{d\epsilon_i\}} \epsilon_i N dp$. The total energy here is thought to be subdivided into a set of intervals $\{d\epsilon_i\}$. The entropy (71) does not depend explicitly on the numbers s . Formula (71) also implies that the wavepackets (as units representing s quanta) are thermodynamically independent.^(246,262) The number of all wavepackets in V and $d\epsilon$, including the empty ones, is given by (61), so we have

$$\sum_{\{s\}} p(s, \epsilon) d\epsilon = g_p. \quad (72)$$

The total number of quanta in V and $d\epsilon$ is Ndp , so

$$\sum_{\{s\}} s p(s, \epsilon) d\epsilon = Ndp. \quad (73)$$

With (70), (73) and Stirling's approximation $p! \approx p \ln p$ the entropy in thermal equilibrium may be written as $S = k \left(bE_{\text{tot}} - cN_{\text{tot}} - \sum_{\{d\epsilon_i\}} g_p \ln [a(\epsilon_i) d\epsilon_i / g_p] \right)$ hence the above-mentioned thermodynamic relations lead to

$$b = 1/kT, \quad c = \mu/kT.$$

So far all mathematical operations could be carried out even if the numbers s, s', r, r' were not integers.

6.4 The Bose and Fermi Distributions

Now we go to special cases. First we take the kind-1 packets to be *Bose* packets. In this case the numbers s, r and n are non-negative integers and (72) becomes

$$\sum_{s=0}^{\infty} p(s, \epsilon) d\epsilon = g_p = \frac{ad\epsilon}{1 - \exp[-(\epsilon - \mu)/kT]}. \quad (74)$$

From this we obtain $ad\epsilon = g_p(1 - \exp[-(\epsilon - \mu)/kT])$ and

$$p_B(s, \epsilon) d\epsilon = g_p (1 - \exp[-(\epsilon - \mu)/kT]) \exp[-s(\epsilon - \mu)/kT]. \quad (75)$$

This formula coincides with Bose's expression for the number of phase space cells occupied with s quanta. The total number of quanta (73) becomes

$$Ndp = \sum_{s=0}^{\infty} s p(s, \epsilon) d\epsilon = \frac{g_p}{\exp[(\epsilon - \mu)/kT] - 1} \quad (76)$$

and we identify these quanta, not the wavepackets, with the particles in the usual interpretation of formula (60). Thus we have arrived at the desired Bose-Einstein distribution function, Eq. (60) with the minus sign.

Let us further consider photons, as a special kind of bosons, and let us consider the processes where photons are absorbed and emitted by atoms. In this case we take the packets of kind 1 to be the photon packets and the packets of kind 2 to be the atoms. In one respect the situation goes beyond the scheme of Fig. 5, in that the number of photons is no longer conserved. In its interaction with the atom the photon is absorbed and exists no longer. Thus, in Fig. 5 the arrow that points from the n -packet in $d\epsilon_1^i$ (second line) to the $(r+n)$ -packet in $d\epsilon_1^f$ (third line) no longer exists, and the $(r+n)$ -packet remains an r -packet. Equivalently, one may say that the n -packet turns into an empty packet ($n = 0$).

Thus, in the balance relation (69) the function $p(r+n, \epsilon_1^f)$ on the right-hand side becomes equal to the function $p(r, \epsilon_1^f)$ on the left-hand side and disappears from

the equation. In the energy-conservation relation (66) we have to put $n = 1$ and $\epsilon_1^f = 0$. What the atom does beyond satisfying the energy conservation in absorbing and emitting a photon is irrelevant. Likewise, in the reverse process the arrow that points from the $(r + n)$ -packet in $d\epsilon_1^f$ (fourth line) to the n -packet in $d\epsilon_1^f$ (fifth line) no longer exists, or equivalently, refers to an empty packet. This does not, however, affect the balance equation, which thus is

$$p(s, \epsilon_1^i) q(s', \epsilon_2^i) q(r', \epsilon_2^f) = p(s - 1, \epsilon_1^i) q(s' - n', \epsilon_2^i) q(r' + n', \epsilon_2^f). \quad (77)$$

The solution is again given by (70), but only if $c = 0$ in $p(s, \epsilon)$, so we have obtained Planck's law.

There is no spontaneous emission in our treatment. Let us compare this with Einstein's treatment. Einstein's balancing equation (Sec. 3 in his 1917 paper⁽¹⁶⁷⁾) is

$$\exp(-\epsilon_n/kT) \rho = \exp(-\epsilon_m/kT) (\rho + A_m^n/B_m^n) \quad (78)$$

where $\rho = h\nu Ndp/(Vd\nu)$ is $h\nu$ times the (time averaged) number of photons per unit volume and per unit frequency interval in the cavity. This equation, unlike our Eq. (69) or (77), is concerned with the number of quanta, not with the number of wavepackets. The second term in the bracket, $A_m^n/B_m^n = 4\pi h\nu^3/c^3$ (polarized radiation), is independent of ρ and is the spontaneous-emission term. One may obtain this term from our Eq. (77) if one puts back the transition probabilities W_1 and W_2 into this equation and uses a special property of them. The left-hand side of (77) effectively means a process where an atom absorbs a photon from an s -photon packet, and the right-hand side a process where an atom emits a photon into an $(s - 1)$ -packet. In order to compare with Einstein's 1917 treatment, which disregards the wavepacket structure of the radiation, one has to sum Eq. (77), with W_1 and W_2 restituted, over all photon packets, i.e. over all values of s :

$$\begin{aligned} q(s', \epsilon_2^i) q(r', \epsilon_2^f) \sum_{s=0}^{\infty} p(s, \epsilon_1^i) W_1(s, \alpha) = \\ = q(s' - n', \epsilon_2^i) q(r' + n', \epsilon_2^f) \sum_{s=0}^{\infty} p(s - 1, \epsilon_1^i) W_2(s - 1, \beta). \end{aligned} \quad (79)$$

α and β are the other arguments in W_1 and W_2 , which do not depend on s . Now we use the special property

$$W_1(s, \alpha) = W_2(s - 1, \beta) = f \cdot s \quad (80)$$

where f may depend on anything but s . This fits with Dirac's statement that the probability of a transition in which a boson is absorbed from (emitted into) state x is proportional to the number of bosons originally in state x (in state x , plus one).⁽²²³⁾

With Eq. (80) one may write (79) in the form

$$\underbrace{\frac{q(s', \epsilon_2^i)}{q(s' - n', \epsilon_2^i)}}_{C_1} d\epsilon_1^i \sum_{s=1}^{\infty} s p(s, \epsilon_1^i) = \underbrace{\frac{q(r' + n', \epsilon_2^f)}{q(r', \epsilon_2^f)}}_{C_2} d\epsilon_1^i \sum_{s=1}^{\infty} s p(s - 1, \epsilon_1^i)$$

$$= C_2 \left[d\epsilon_1^i \sum_{s=1}^{\infty} (s - 1) p(s - 1, \epsilon_1^i) + d\epsilon_1^i \sum_{s=1}^{\infty} p(s - 1, \epsilon_1^i) \right],$$

and with (74) and (76) one obtains $C_1 Ndp = C_2 (Ndp + g_p)$. If one multiplies this by $h\nu/(Vd\nu)$ and observes that $C_1/C_2 = \exp[n'(\epsilon_2^f - \epsilon_2^i)/kT]$ and $n' = 1$ one obtains Einstein's balancing equation (78). We note that it is only the special form (78) of the balancing equation that requires the special property (80) of the transition probabilities in order to arrive at the Planck distribution.

Second, we take the kind-1 packets to be *Fermi* packets. In this case we have only zero- and one-quantum packets, and the sums in (74) and (76) range only from 0 to 1. Thus

$$\sum_{s=0}^1 p(s, \epsilon) d\epsilon = ad\epsilon(1 + \exp[-(\epsilon - \mu)/kT]) = g_p.$$

Hence

$$p_F(s, \epsilon) d\epsilon = g_p (1 + \exp[-(\epsilon - \mu)/kT])^{-1} \exp[-s(\epsilon - \mu)/kT] \quad (81)$$

and the total number of quanta is

$$Ndp = \sum_{s=0}^1 s p(s, \epsilon) d\epsilon = \frac{g_p}{\exp[(\epsilon - \mu)/kT] + 1}$$

which is the desired Fermi-Dirac distribution, i.e. formula (60) with the plus sign.

The balance relation (69) in the Fermi case may be specified to read ($s = n = 1, r = 0$)

$$p(1, \epsilon_1^i) p(0, \epsilon_1^f) q(s', \epsilon_2^i) q(r', \epsilon_2^f) = p(0, \epsilon_1^i) p(1, \epsilon_1^f) q(s' - n', \epsilon_2^i) q(r' + n', \epsilon_2^f)$$

and may also be given a specific interpretation: a fermi 1-packet in $d\epsilon_1^i$ is altered in an interaction with a wavepacket of kind 2 whereby it goes into $d\epsilon_1^f$, conservation of energy requiring $\epsilon_1^i - \epsilon_1^f = n'(\epsilon_2^f - \epsilon_2^i)$. Then it goes into a region of phase space within $d\epsilon_1^f$ that is not yet occupied by a (non-empty) wavepacket. The transition rate is proportional to the size of this region, expressed in fundamental units, that is, to the number of empty packets in $d\epsilon_1^f$, $p(0, \epsilon_1^f)d\epsilon_1^f$.

6.5 Quantum Count Fluctuations

Finally we extend our considerations on wavepackets in a cavity to include fluctuations. To be definite, we consider a small subvolume v of the total cavity volume V . We imagine that the subvolume is homogeneously filled with detectors (groups of sensitive atoms). The detectors are sensitive only within the interval $p \dots p + dp$ of the absolute value of momentum or the corresponding energy interval $\epsilon \dots \epsilon + d\epsilon$, and we assume that within this interval the sensitivity is constant. These detectors are switched on during the interval Δt , and the number of counts is registered. This procedure is repeated a great many times, where the time intervals between the repetitions are large compared with Δt . We then ask for the mean square deviation, or variance, $(\Delta m)^2$ of the number of counts. Again we shall treat both boson and fermion wavepackets on an equal footing.

The subvolume v together with the interval dp define a certain volume of phase space and with this a certain number g_v of (empty plus non-empty) wavepackets, given by formula (61) with V replaced by v . The number of wavepackets with which the counter can interact during Δt is larger than g_v because (1) the switch-on time Δt may be so long that many sets of wavepackets, each set filling the counter volume at one time, may pass through the counter during Δt , and (2) the counter can also interact with wavepackets that only partially extend into it. The (integer) number of wavepackets that partially and/or totally cover the phase-space region of the counter during Δt is denoted by g , where $g \geq g_v$ and $g \geq 1$. A count is always an interaction of the counter with a wavepacket, not with a quantum. The number of quanta represented by the g packets fluctuates because between two measurements some few-quantum packets may have replaced some many-quantum packets and vice versa. This is the only source of fluctuations. Fluctuations that arise from a non-empty wavepacket leaving the region without another non-empty packet entering it are already accounted for because our number of wavepackets includes empty packets, so that a non-empty packet leaving the region is equivalent with an empty packet entering it.

Of all the quanta of a wavepacket only some fraction will be counted because the counter has a limited intrinsic efficiency, is switched on only during Δt , and may cover only part of the wavepacket. The probability of m counts from an n -quantum wavepacket can be shown to be the binomial distribution function

$$b(m; n, \eta) = \binom{n}{m} \eta^m (1 - \eta)^{n-m} \quad (82)$$

where $\eta = \bar{m}/n$ ($0 \leq \eta \leq 1$) is the average fraction of quanta of the packet that are counted during the interval Δt . Such a formula has already been derived in Sec. 3.2 [Formula (24)] with the condensed packet being described by the product wave function

$$\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N, t) = \varphi(\mathbf{x}_1, t) \varphi(\mathbf{x}_2, t) \dots \varphi(\mathbf{x}_N, t). \quad (83)$$

In Sec. 3.2, however, we were asking for the probability of counting the quanta in a finite region D^3 of space, and we assumed there, in accordance with standard

quantum mechanics, that we would count all of them with certainty in the entire space R^3 during Δt , that is $\int_{-\infty}^{+\infty} |\Psi|^2 d^3x_1 \dots d^3x_N = 1$. Therefore we would have obtained $\eta = 1$ if $D^3 = R^3$. Now, the detector volume v may happen to be much larger than a wavepacket, and in this case v is effectively equivalent to R^3 , leading to $\eta = 1$. Formula (82) would then always give zero for $m \neq n$ and be useless. It is, nevertheless, possible to maintain formula (82) even in this situation if we amplify the meaning of η . In fact, in our interpretation the probability of a wavepacket inducing an effect (count) anywhere in space during Δt may be admitted to be less than unity: according to formula (40) of Sec. 4.2 the probability of the wavepacket $\psi(\mathbf{x}, t)$ inducing an effect in d^3x during Δt may be written as

$$W_1 = \kappa(\Delta t) |\psi(\mathbf{x}, t)|^2 d^3x,$$

so that the probability that ψ induces an effect anywhere in space during Δt is, by (41),

$$I'_1 = \kappa(\Delta t) \int_{-\infty}^{+\infty} |\psi(\mathbf{x}, t)|^2 d^3x = \kappa(\Delta t) \leq 1,$$

with ψ still being normalized, $\int_{-\infty}^{+\infty} |\psi|^2 d^3x = 1$. Now instead of P_1 given by Formula (22) of Sec. 3.2 we use

$$W_1 = \kappa(\Delta t) \int_{\mathbf{x} \in D^3} |\varphi(\mathbf{x}, t)|^2 d^3x$$

and, observing that all quanta are to be treated in the same way, instead of P_2 in (23) we use

$$W_2 = \binom{N}{m} \left(\kappa \int_{D^3} |\varphi|^2 d^3x \right)^m \left(1 - \kappa \int_{D^3} |\varphi|^2 d^3x \right)^{N-m},$$

which would be equal to P_2 if $\kappa = 1$. If we now, instead of (25), define η as

$$\eta := \kappa(\Delta t) \int_{D^3} |\varphi(\mathbf{x}, t)|^2 d^3x \quad (84)$$

we can maintain formulas (24) and (82), independently of whether the counter volume covers the wavepacket totally or partially. The difference between these two cases is absorbed in the numerical value of η , η suffering an additional decrease when we go from the case of total to that of partial spatial covering.

What, then, is the probability $W(m; g)$ of counting m quanta from g wavepackets? To answer this question we first evaluate the probability $w(n; g)$ that the g packets represent n quanta and then the probability $B(m; n)$ that of these n quanta m are counted, and we write $W(m; g) = \sum_{n=m}^{\infty} w(n; g) B(m; n)$.

We first consider *boson* packets. The probability that a randomly chosen boson packet is an s -quantum packet is given by (75), which we write in the form

$$p(s) d\epsilon / g_p = \frac{1}{(1 + \bar{s})} \frac{1}{(1 + 1/\bar{s})^s}. \quad (85)$$

\bar{s} is the average number of quanta represented by one packet averaged over all packets in V and dp (or the corresponding $d\epsilon$), $\bar{s} = \sum_{s=0}^{\infty} s p(s) d\epsilon / g_p$.

The probability that of the g packets in v the first one represents s_1 , the second s_2, \dots and the g -th s_g quanta, with $\sum_{i=1}^g s_i = n$, is the product

$$\prod_{i=1}^g p(s_i) d\epsilon / g_p = \frac{1}{(1 + \bar{s})^g} \frac{1}{(1 + 1/\bar{s})^n}. \quad (86)$$

We are, however, not interested in the particular distribution (which quanta are represented by which packets), so we have to form a sum of expressions (86), one for each distribution. Since (86) is the same for any distribution we need only to multiply (86) by the number of possible distributions, given by the well-known combinatorial expression

$$\binom{g+n-1}{n} \equiv \frac{(g+n-1)!}{(g-1)! n!}. \quad (87)$$

Thus the probability that the g packets represent n quanta is

$$w(n; g) = \binom{g+n-1}{n} \frac{1}{(1 + \bar{s})^g} \frac{1}{(1 + 1/\bar{s})^n}. \quad (88)$$

This formula was already obtained by Mandel⁽²⁶³⁾ in a related context.

Next we evaluate the probability $B(m; n)$ that if n quanta are present m are counted. Here we take advantage of the fact that this probability is independent of how these quanta are represented by the different wavepackets. This is seen in the following way: assume that all the n quanta are from one and the same wavepacket. Then the probability of m counts is given by the binominal distribution (82). Next assume that the quanta are from two wavepackets, one with n_1 and one with $n_2 = n - n_1$ quanta. The probability of m counts would then be

$$\sum_{\underbrace{m_1 \quad m_2}_{m_1+m_2=m}} \binom{n_1}{m_1} \eta^{m_1} (1 - \eta)^{n_1-m_1} \binom{n_2}{m_2} \eta^{m_2} (1 - \eta)^{n_2-m_2}.$$

But due to a special folding property of the binominal distribution^(264,265) this is equal to $b(m; n, \eta)$ of (82). And this would remain so if the quanta were from any number of packets. We may thus assume that all n quanta are from one and the same wavepacket whence $B(m; n) = b(m; n, \eta)$ of (82). In deriving this result we have used the same value of η for all wavepackets. This requires a justification because when the phase-space volume of a wavepacket $\Delta^3 x \Delta^3 p$ and that of the counter $4\pi v p^2 dp$ with which it interacts are comparable, their overlap and with this the value of η may vary appreciably from one wavepacket to the next, even if, as we assume, all wavepackets in the cavity have very nearly the same size in

phase space. In this case the above folding theorem in fact no longer holds generally, although still in special cases ($\eta \ll 1, \eta \approx 1$, Poisson approximation, normal approximation⁽²⁶⁶⁾). When, however, the phase-space volume of the counter is large (in each direction) compared with that of a wavepacket the counter covers almost each wavepacket completely, and since we consider a counter with constant sensitivity over its whole phase-space volume, η is still the same for all packets. When, in the opposite limit, the counter is small compared with the wavepacket, it is true that η may vary considerably because the counter may cover regions of the wavepacket with varying $|\psi|^2$ (in x space or in p space) since we have not assumed that $|\psi|^2$ is constant over the wavepacket. But now, whichever region of a wavepacket is covered by the counter, η will always be small, and under this condition the folding theorem still holds in the form

$$\underbrace{\sum_{m_1} \sum_{m_2}}_{m_1+m_2=m} b(m_1; n_1, \eta_1) b(m_2; n_2, \eta_2) - b\left(m; n_1 + n_2, \frac{n_1 \eta_1 + n_2 \eta_2}{n_1 + n_2}\right) \\ \propto (\eta_1 - \eta_2)^2 + \text{terms of higher order in } \eta_1 \text{ and } \eta_2 ,$$

so that $B(m; n) = b(m; n, \eta)$ still holds with η representing some average over different regions of the wavepacket. Therefore we consider $B(m; n) = b(m; n, \eta)$ as an acceptable approximation.

With this the probability $W(m; g)$ of counting m quanta when g packets are in $4\pi\nu p^2 dp$ is

$$W(m; g) = \sum_{n=m}^{\infty} w(n; g) b(n; m, \eta) = \sum_{n=m}^{\infty} w(n; g) \binom{n}{m} \eta^m (1 - \eta)^{n-m} \quad (89) \\ = \sum_{n=m}^{\infty} \binom{g+n-1}{n} \frac{1}{(1 + \bar{s})^g} \frac{1}{(1 + 1/\bar{s})^n} \binom{n}{m} \eta^m (1 - \eta)^{n-m}.$$

Formula (89) is the same as that obtained via the standard quantization formalism,⁽²⁶⁷⁾ as it should be. Substituting $l = n - m$ and using the binomial identity $\binom{l+g-1+m}{l} = (-1)^l \binom{-g-m}{l}$ and Newton's binomial formula $\sum_{l=0}^{\infty} \binom{-g-m}{l} \left(-\frac{1-\eta}{1+1/\bar{s}}\right)^l = \left(1 - \frac{1-\eta}{1+1/\bar{s}}\right)^{-g-m}$ we obtain

$$W(m; g) = \binom{g+m-1}{m} \frac{1}{(1 + \eta \bar{s})^g} \frac{1}{(1 + 1/(\eta \bar{s}))^m}. \quad (90)$$

From this we can calculate the variance $(\Delta m)^2 := \sum_{m=0}^{\infty} (m - \bar{m})^2 W(m; g) = \eta \bar{s} g (1 + \eta \bar{s})$ [use the generating function⁽²⁶⁸⁾ of $W(m; g)$]. Observing that $\eta \bar{s} g = \bar{m}$ is the average number of counted quanta from g packets we may finally write the variance of the number m of counts as

$$(\Delta m)^2 = \bar{m} \left(1 + \frac{\bar{m}}{g}\right). \quad (91)$$

Comparing (90) with (88) we see that the distribution, and hence variance, of the counted quanta has the same form as that of the existing quanta.

Our procedure is easily carried over to the *Fermi* case. We only have to observe that there are only zero- and one-quantum packets. Of course (83) no longer holds, but (82) still does since for $m \leq n \leq 1$ it reduces to three trivial expressions. Further, instead of (85) we have to take (81) which we write in the form

$$p(s)d\epsilon/g_p = (1 - \bar{s}) \frac{1}{(1/\bar{s} - 1)^s}, \quad (92)$$

and instead of (87) we have to take

$$\binom{g}{n} = \frac{g!}{(g-n)!n!}. \quad (93)$$

(90) is then replaced by

$$W(m; g) = \binom{g}{m} (1 - \eta\bar{s})^g \frac{1}{(1/(\eta\bar{s}) - 1)^m} \quad (94)$$

and (91) by

$$(\Delta m)^2 = \bar{m} \left(1 - \frac{\bar{m}}{g} \right). \quad (95)$$

For $g = 1$ Formula (90) becomes ($\bar{m} = \eta\bar{s}g$)

$$W_B(m; 1) = \frac{\bar{m}^m}{(1 + \bar{m})^{m+1}} \quad (\text{Bose}) \quad (96)$$

and (94)

$$W_F(m; 1) = \frac{\bar{m}^m}{(1 - \bar{m})^{m-1}} \quad (\text{Fermi}). \quad (97)$$

For $g \rightarrow \infty$ both (90) and (94) approach the Poisson distribution

$$W_{BM}(m; \infty) = \frac{\bar{m}^m}{m!} e^{-\bar{m}} \quad (\text{Boltzmann}). \quad (98)$$

It seems that our results not only hold for cavities but also for beams, at least in some situations. This we infer from the observation that the distributions (96) and (98) are also obtained for stationary beams of chaotic light in standard quantized radiation theory⁽²⁶⁹⁾ and are confirmed by experiment.⁽²⁷⁰⁾ Formula (96) obtains in the limit when the counting time Δt is short compared with the coherence time τ_c , $\Delta t \ll \tau_c$, and (98) obtains in the opposite limit, $\Delta t \gg \tau_c$. These limits can be compared with the limits $g = 1$ and $g \gg 1$ of our treatment because the wavepacket structure of the radiation field reflects its coherence properties, in that the spatial size of a photon wavepacket is a measure of the size of the coherence region with length $L_c = c\tau_c$ (Appendix A). The limit $\Delta t \gg \tau_c$ means that many

wavepackets can interact with the counter during Δt , either consecutively, if the packets are large, or simultaneously, if they are small compared with the counter volume v . Thus $g \gg 1$ in this limit. In the opposite limit, $\Delta t \ll \tau_c$, the wavepackets can interact with the counter only during such a short time interval that their movement is negligible. Whether we have $g \gg 1$ or $g = 1$ now depends on further specifications. When the packets are large and pass over the counter one after the other we have $g = 1$. This was in fact implicitly assumed in the (plane-wave) calculations referred to by Loudon⁽²⁶⁹⁾ and was explicitly stated in the experimental verification⁽²⁷⁰⁾ of Formula (96). So, here $\Delta t \gg \tau_c$ means $g \gg 1$, and $\Delta t \ll \tau_c$ means $g = 1$, and the situations are those covered by our formulas.

Acknowledgement

I thank my colleagues Professors Júlio de Melo Teixeira and Nelson Lima Teixeira for taking over part of my teaching load in a difficult phase, Mr. W. Peter Brown, Bielefeld, for looking over the English, Professor Dr. Harvey R. Brown, Oxford, for a critical reading of the manuscript and stimulating discussions, and Professor Dr. Georg Süßmann and the Physics Section of the University of Munich for their hospitality.

APPENDIX A: COLLECTION OF WAVEPACKET SPREADING FORMULAS

The wavepackets considered here are the usual free packets of Schrödinger or de Broglie waves. Since the formulas are rather spread out in the literature and are often restricted to special cases we have collected here some more general results for easy reference.

The general wavepacket is written in the form

$$\psi(\mathbf{x}, t) = (2\pi)^{-3/2} \int_{-\infty}^{+\infty} \tilde{\psi}(\mathbf{k}) \exp[i(\mathbf{k}\mathbf{x} - \omega(\mathbf{k})t)] d^3k \quad (\text{A1})$$

where $\tilde{\psi}(\mathbf{k})$, and hence $\psi(\mathbf{x}, t)$, is normalized

$$\int_{-\infty}^{+\infty} |\tilde{\psi}(\mathbf{k})|^2 d^3k = 1. \quad (\text{A2})$$

The (three-dimensional) Fourier transform of $\psi(\mathbf{x}, t)$ is

$$\tilde{\psi}(\mathbf{k}, t) = \tilde{\psi}(\mathbf{k}) \exp[-i\omega(\mathbf{k})t]. \quad (\text{A3})$$

This is the wavepacket in k (momentum) space. Mathematically the wavepacket need not have a sharp boundary in x space (or in k space), but for practical purposes it may be considered to have a finite extension given, for example, by the standard deviation

$$\Delta x := \langle (x - \langle x \rangle)^2 \rangle^{1/2} \quad (\text{A4})$$

with

$$\langle x \rangle(t) = \int_{-\infty}^{+\infty} \psi^*(\mathbf{x}, t) x \psi(\mathbf{x}, t) d^3x, \quad (\text{A5})$$

and with analogous expressions for the widths Δy and Δz . In the special case of the Gaussian form (and $t = 0$)

$$|\psi(x)|^2 = (2\pi\sigma^2)^{-1/2} \exp[-(x - \langle x \rangle)^2 / (2\sigma^2)] \quad (\text{A6})$$

Δx of (A4) is equal to the parameter σ and is the distance between the maximum and the point where the distribution has fallen off to $\exp(-1/2) = 0.61$ of the maximum value. Sometimes the form of $\psi(\mathbf{x})$ or of $\tilde{\psi}(\mathbf{k})$ is such that the integrals in (A4) etc. diverge, for example for $\tilde{\psi}(\mathbf{k}) \propto \sin(ak)/k$ and for $\tilde{\psi}(\mathbf{k}) \propto (1 + a^2 k^2)^{-1/2}$. In such cases one uses other definitions of the width, for instance the distance between the maximum and the first zero, the half width, the equivalent width,⁽²⁷¹⁾ or the overall width.⁽²⁷²⁾

The dispersion law $\omega(\mathbf{k})$ is determined by the relation between the energy E and the momentum \mathbf{p} of the object the wavepacket is to represent provided we make use of the Einstein-Planck relation

$$\nu = E/h \quad \text{or} \quad \omega = E/\hbar \quad (\text{A7})$$

and the de Broglie relation

$$\mathbf{p} = \hbar \mathbf{k} \quad (\text{A8})$$

where

$$k \equiv |\mathbf{k}| = 2\pi/\lambda. \quad (\text{A9})$$

Thus, the relativistic relation for a free particle

$$E = E_{\text{tot}} = \pm(p^2 c^2 + m^2 c^4)^{1/2} \quad (\text{A10})$$

leads to

$$\omega(\mathbf{k}) = \pm c(k^2 + \kappa^2)^{1/2} \quad (\text{A11})$$

where

$$\lambda_C = 1/\kappa = \hbar/(mc) \quad (\text{A12})$$

is the Compton length belonging to the rest-mass parameter m .

The most general solution of a relativistic wave equation would be a function of the type (A1) in which $\omega(\mathbf{k})$ from (A11) has the positive sign plus a function (A1) in which $\omega(\mathbf{k})$ has the negative sign. That is, the most general form would be a superposition of waves with positive as well as those with negative frequencies viz. energies. Although we shall consider the general relativistic formula (A11) we will restrict ourselves to positive-energy wave functions in order to have a simple connection with the results of nonrelativistic quantum mechanics.

Let us consider the time dependence of the wavepacket's "center" $\langle \mathbf{x} \rangle$ and "width"

$$\sigma(t) = ([\Delta x(t)]^2 + [\Delta y(t)]^2 + [\Delta z(t)]^2)^{1/2} \quad (\text{A13})$$

where

$$[\Delta x(t)]^2 := \int (x - \langle x \rangle)^2 |\psi(\mathbf{x}, t)|^2 d^3x \quad (\text{A14})$$

and analogously for the y and z components. A very general calculation of these quantities has been given by Bradford.⁽²⁷³⁾ His treatment works in three dimensions and does not require a special form either for $\tilde{\psi}(\mathbf{k})$ or for $\omega(\mathbf{k})$, except for the usual convergence requirements of the integrals that appear in the averaging procedures. In particular, the treatment is valid in the nonrelativistic as well as in the relativistic domain. The result for the center is

$$\langle \mathbf{x} \rangle(t) = \langle \mathbf{x} \rangle(0) + \langle \vec{v}_g \rangle t \quad (\text{A15})$$

with

$$\langle \mathbf{x} \rangle(0) = - \int \text{Im} \left\{ \tilde{\psi}^* \nabla_k \tilde{\psi} \right\} d^3k \quad (\text{A16})$$

$$\langle \vec{v}_g \rangle = \int |\tilde{\psi}|^2 \nabla_k \omega d^3k \quad (\text{A17})$$

$$\tilde{\psi} = \tilde{\psi}(\mathbf{k}).$$

According to (A15) and (A17) the center moves at constant velocity $\langle \vec{v}_g \rangle$ which is the mean group velocity.

The result for the width is

$$\sigma^2(t) = \sigma^2(t_0) + [\Delta v_g]^2 (t - t_0)^2 \quad (\text{A18})$$

where

$$\sigma^2(t_0) = \sigma^2(0) - [\Delta v_g]^2 t_0^2 \quad (\text{A19})$$

$$\sigma^2(0) = - \int \text{Re} \left\{ \tilde{\psi}^* \nabla_k^2 \tilde{\psi} \right\} d^3 k - [\langle \mathbf{x} \rangle(0)]^2 \quad (\text{A.20})$$

$$t_0 = \left(\int \text{Im} \left\{ \tilde{\psi}^* \nabla_k \tilde{\psi} \right\} \nabla_k \omega d^3 k + \langle \mathbf{x} \rangle(0) \langle \vec{v}_g \rangle \right) [\Delta v_g]^{-2} \quad (\text{A21})$$

and

$$[\Delta v_g]^2 := \int |\tilde{\psi}|^2 (\nabla_k \omega - \langle \vec{v}_g \rangle)^2 d^3 k. \quad (\text{A22})$$

Formula (A18) shows that the width varies hyperbolically with time. This type of variation is even independent of the form of the dispersion law $\omega(\mathbf{k})$, except when Δv_g is zero: there is either a hyperbolic dependence or none. The minimum extension of the wavepacket occurs at $t = t_0$; before t_0 the wavepacket contracts, after t_0 it spreads out.

When we write $\tilde{\psi}(\mathbf{k})$ in the polar form

$$\tilde{\psi}(\mathbf{k}) = \rho(\mathbf{k}) \exp[i\alpha(\mathbf{k})] \quad (\text{A23})$$

we are led to

$$\langle \mathbf{x} \rangle(0) = -\langle \nabla_k \alpha \rangle \quad (\text{A24})$$

$$t_0 = (\langle \nabla_k \alpha \nabla_k \omega \rangle - \langle \nabla_k \alpha \rangle \langle \nabla_k \omega \rangle) [\Delta v_g]^{-2}. \quad (\text{A25})$$

We thus can fix the time and the place of the minimum extension by an appropriate choice of the phase $\alpha(\mathbf{k})$, that is, by $\langle \nabla_k \alpha \rangle$ and $\langle \nabla_k \alpha \nabla_k \omega \rangle$. Resolving the two Eqs. (A24) and (A25) for these two quantities we obtain

$$\langle \nabla_k \alpha \rangle = \langle \nabla_k \omega \rangle t_0 - \langle \mathbf{x} \rangle(t_0) \quad (\text{A26})$$

$$\langle \nabla_k \alpha \nabla_k \omega \rangle = \langle [\nabla_k \omega]^2 \rangle t_0 - \langle \nabla_k \omega \rangle \langle \mathbf{x} \rangle(t_0). \quad (\text{A27})$$

For example, the especially simple form with three parameters $\vec{\xi}_0, \tau_0$ and η_0

$$\alpha(\mathbf{k}) = -\mathbf{k} \vec{\xi}_0 + \omega(\mathbf{k}) \tau_0 + \eta_0 \quad (\text{A28})$$

leads to

$$\langle \nabla_k \alpha \rangle = -\vec{\xi}_0 + \tau_0 \langle \nabla_k \omega \rangle \quad (\text{A29})$$

which, by comparison with (A26), shows that the parameters $\vec{\xi}_0$ and τ_0 coincide with the initial values

$$\vec{\xi}_0 = \langle \mathbf{x} \rangle(t_0), \quad \tau_0 = t_0, \quad (\text{A30})$$

and η_0 is an arbitrary constant.

In what follows we will always assume (A28) with $\vec{\xi}_0=0$ and $\tau_0=0$, so that the minimum extension occurs at $t = t_0 = 0$, and the center of the packet at that time is $\langle \mathbf{x} \rangle(0) = 0$. Formulas (A18) and (A20) then simplify to

$$\sigma^2(t) = \sigma^2(0) + [\Delta v_g]^2 t^2 \quad (\text{A31})$$

and

$$\sigma^2(0) = - \int \text{Re} \left\{ \tilde{\psi}^* \nabla_k^2 \tilde{\psi} \right\} d^3 k. \quad (\text{A32})$$

Let us now consider the spreading velocity

$$v_s := \partial \sigma(t) / \partial t = [\Delta v_g]^2 t / \sigma(t) \quad (\text{A33})$$

which for large t tends to the asymptotic spreading velocity

$$v_{s\infty} := \lim_{t \rightarrow \infty} v_s = \Delta v_g. \quad (\text{A34})$$

The packet spreads out to double its initial ($t = 0$) extension in a time $\tau_2 = \sqrt{3}\sigma(0)/\Delta v_g$. After this time the spreading velocity is $v_s = (\sqrt{3}/2)\Delta v_g = 0.87v_{s\infty}$. That is, 87% of the asymptotic spreading velocity is already reached when the packet has doubled its initial extension. The asymptotic value may thus be used in all practical estimates. At the time t_1 when the asymptotic spreading velocity has (practically) been reached, the time for the wavepacket to further double its extension is $\sigma(t_1)/\Delta v_g$ which is somewhat smaller than τ_2 provided we identify $\sigma(t_1)$ with $\sigma(0)$ in this comparison.

To proceed further we must make specific assumptions about $\tilde{\psi}(\mathbf{k})$. We shall assume a nearly unidirectional and quasimonochromatic packet, that is, a *narrow packet* in k space; that is, $\tilde{\psi}(\mathbf{k})$ is assumed to be appreciably different from zero only in a narrow region concentrated about the point $\mathbf{k}_0 = (0, k_0, 0)$ so that

$$\Delta k_x, \Delta k_y, \Delta k_z \ll |\mathbf{k}_0| \equiv k_0. \quad (\text{A35})$$

With the help of the Fourier reciprocity (Heisenberg) relations

$$\Delta x(0) \Delta k_x \geq 1/2 \quad \text{etc.}, \quad (\text{A36})$$

where the equality sign can only occur for a Gaussian form for $\tilde{\psi}(\mathbf{k})$, the condition (A35) can be written as a condition in ordinary space

$$\Delta x(0), \Delta y(0), \Delta z(0) \gg 1/(2k_0). \quad (\text{A37})$$

It is then possible to express Δv_g as a function of Δk_x and Δk_y . In the case where $k_0 \neq 0$ it is reasonable to consider separately the longitudinal spreading, along the direction y of the center, and the transverse spreading, normal to that direction, say in the x direction. We then expand $\nabla_k \omega = c^2 \mathbf{k} / \omega(\mathbf{k})$ in a three-dimensional Taylor series about \mathbf{k}_0 and break the series off after the quadratic terms. After a straightforward but tedious calculation one arrives at

$$\begin{aligned} v_{sx\infty} &:= \lim_{t \rightarrow \infty} \partial \Delta x(t) / \partial t \\ &= \Delta v_{gx} := \left(\int |\tilde{\psi}|^2 (\partial \omega / \partial k_x - \langle v_{gx} \rangle)^2 d^3 k \right)^{1/2} \\ &= \frac{c^2}{\omega_0} \Delta k_x = \frac{c}{(k_0^2 + \kappa^2)^{1/2}} \Delta k_x \end{aligned} \quad (\text{A38})$$

$$\geq \frac{c}{2(k_0^2 + \kappa^2)^{1/2} \Delta x(0)} \quad (\text{A39})$$

where (A36) has been used for obtaining (A39). Likewise we obtain

$$\vec{v}_0 \equiv \langle \vec{v}_g \rangle = (0, k_0 c^2 / \omega_0, 0) \quad (\text{A40})$$

$$v_0 = k_0 c^2 / \omega_0, \quad \omega_0 = c(k_0^2 + \kappa^2)^{1/2}, \quad (\text{A41})$$

and with this we may write (A38) as

$$v_{sx\infty} = \frac{c}{\kappa} \Delta k_x \left(1 - (v_0/c)^2 \right)^{1/2} \quad (\text{A42})$$

$$\geq \frac{c \left(1 - (v_0/c)^2 \right)^{1/2}}{2\kappa \Delta x(0)}. \quad (\text{A43})$$

In the same way we arrive at the longitudinal asymptotic spreading velocity

$$v_{sy\infty} = \left(\frac{c\kappa}{\omega_0} \right)^2 \frac{c^2}{\omega_0} \Delta k_y = \frac{c}{\kappa} \Delta k_y \left(1 - (v_0/c)^2 \right)^{3/2} \quad (\text{A44})$$

$$\geq \frac{c \left(1 - (v_0/c)^2 \right)^{3/2}}{2\kappa \Delta y(0)} \quad (\text{A45})$$

and

$$\begin{aligned} v_{sy\infty} / v_0 &= \Delta k_y / k_0 \times (1 + (k_0/\kappa)^2)^{-1} \\ &= \Delta k_y / k_0 \times (1 - (v_0/c)^2). \end{aligned} \quad (\text{A46})$$

We thus have the interesting result that the spreading velocities of a narrow packet in k space do not depend on the detailed form of the wavepacket over and above

its second central moments $\Delta k_x, \Delta k_y$. In particular, formulas (A38) to (A46) hold in the nonrelativistic as well as in the relativistic domain.

From (A43) it might appear that $v_{sx\infty} \rightarrow \infty$ if $\Delta x(0) \rightarrow 0$. This is not true because by (A36) $\Delta x(0) \rightarrow 0$ would imply $\Delta k_x \rightarrow \infty$, and our assumption of a narrow wavepacket would no longer hold. In fact, closer inspection shows that the spreading velocity (A33) is always limited when $\Delta x(0) \rightarrow 0$.⁽²⁷⁴⁾ For equal widths $\Delta k_x = \Delta k_y$ the transverse spreading is always larger than the longitudinal one since by (A42) and (A44) it is then

$$\frac{v_{sx\infty}}{v_{sy\infty}} = \frac{1}{1 - (v_0/c)^2} \geq 1. \quad (\text{A47})$$

In the *nonrelativistic domain* we have

$$k_0^2 \ll \kappa^2, \quad (\text{A48})$$

and condition (A35) now takes the form

$$\Delta k_x, \Delta k_y, \Delta k_z \ll \kappa. \quad (\text{A49})$$

The integrals involved in the averaging procedures may then be restricted to regions where the dispersion law (A11) can be approximated by $\omega(\mathbf{k}) = mc^2/\hbar + \hbar k^2/(2m)$ from which one obtains the nonrelativistic relation $\Delta \vec{v}_g = \hbar \Delta \mathbf{k}/m = \Delta \mathbf{p}/m$. With (A36) condition (A49) in ordinary space reads

$$\Delta x(0), \Delta y(0), \Delta z(0) \gg 1/(2\kappa) = \frac{1}{2} \lambda_C. \quad (\text{A50})$$

In the *relativistic domain* the spreading becomes slower as the velocity v_0 of the packet (center) approaches c . In the zero-mass limit, $\kappa \rightarrow 0$, $\omega_0 \rightarrow ck_0$, formulas (A38) and (A44) lead to

$$v_{sx\infty} \rightarrow \frac{c}{k_0} \Delta k_x \quad (\text{A51})$$

$$v_{sy\infty} \rightarrow 0. \quad (\text{A52})$$

Thus for photons there is a finite transverse spreading but no longitudinal spreading, in accordance with the well known absence of spreading of one-dimensional electromagnetic pulses composed of unidirectional waves.

Let us finally consider another length, the *coherence length* $\Delta_c y$ of a wavepacket (in the y direction). We want to show that this length is equal to the coherence length of the beam in which the wavepackets take part and that $\Delta_c y$ does not spread out in time.

The coherence length of a *wavepacket* is defined by means of the autocorrelation function

$$\gamma(b) = \int \psi^*(x, y, z, t) \psi(x, y + b, z, t) d^3x. \quad (\text{A53})$$

The function $|\gamma(b)|$ is maximal ($=1$) at $b = 0$, and $\Delta_c y$ is defined as that value of b where $|\gamma(b)|$ has decayed for the first time to $\exp(-1/2)$. $\Delta_c y$ is closely related to the “mean peak width” of Hilgevoord and Uffink.^(272,275)

The coherence length L_c of a *beam* is defined by means of the “contrast” or “visibility”

$$V(b) = (I_{\max} - I_{\min}) / (I_{\max} + I_{\min}) \quad (\text{A54})$$

where I_{\max} (I_{\min}) is the maximal (minimal) intensity in the interference pattern obtained by dividing the beam into two sub-beams, delaying the one sub-beam by the distance b , and then reuniting the two. Length L_c is defined as that value of b where $V(b)$ has decayed to $\exp(-1/2)$. It is a quantity that is easy to measure.

In order to see that $\Delta_c y$ equals L_c suppose that the beam is a stream of equal wavepackets and that the subdivision of the beam means the subdivision of each single wavepacket. The reunited beam then means reunited wavepackets and for each wavepacket we have $\psi_f = \psi(x, y + b, z, t) + \psi(x, y, z, t)$. The probability that it will cause a count in the final counter, at any moment of time and at any place within the large counter (assuming 100% detection efficiency) is proportional to

$$\begin{aligned} W &= \int_{-\infty}^{+\infty} |\psi_f|^2 d^3x \\ &= \int |\psi(x, y, z, t)|^2 d^3x + \int |\psi(x, y + b, z, t)|^2 d^3x \\ &\quad + 2\text{Re} \int \psi^*(x, y, z, t) \psi(x, y + b, z, t) d^3x. \end{aligned} \quad (\text{A55})$$

The function ψ_f need not be normalized because the original beam may be divided into more than the two sub-beams considered. The integrals in the second line of (A55) are equal and we denote them by A . The third line may then be written as $2A\text{Re}\gamma(b)$ when we use (A53) and observe that ψ in (A53) is normalized but in (A55) perhaps not. Writing $\text{Re}\gamma = |\gamma| \cos \alpha$ we have

$$W(b) = 2A + 2A|\gamma(b)| \cos(\alpha(b)). \quad (\text{A56})$$

Now, the detection probability $W(b)$ of the wavepacket is proportional to the registered time averaged intensity $I(b)$ of the final beam, and when one assumes that $\cos(\alpha(b))$ in typical cases varies much more rapidly than does $|\gamma(b)|$ one obtains $V(b) = |\gamma(b)|$ and with this $\Delta_c y = L_c$.

$\gamma(b)$ and with it $\Delta_c y$ is independent of time because $\gamma(b)$ is just the mean value of the y -translation operator in the state $\psi(x, y, z, t)$, and this operator commutes with the Hamilton operator for free packets.⁽²⁷⁶⁾ Thus, whereas the length $\Delta y(t)$ of the packet may spread out with time, the coherence length $\Delta_c y$ does not and is usually proportional to the minimum length $\Delta y_{\min} = \Delta y(0)$ (assumed to occur at $t = 0$). For a Gaussian packet, for example, one obtains $\Delta_c y = 2\Delta y_{\min}$.⁽²⁷⁷⁾

There is no strict relation between the coherence length $\Delta_c y$ and the length $\Delta y(t)$ of the wavepacket. Any beam, with long or with short coherence length,

may be considered to be one very long wavepacket, simply by superposing the shorter wavepackets that originally were conceived as its constituents. All that can be said is that the coherence length is of the order of a lower bound for the length of the constituent wavepackets. The difference between $\Delta y(t)$ and $\Delta_c y$ may always be small in the case of photon wavepackets which propagate in y direction, since photon packets do not spread out in longitudinal but only in transverse directions. In fact, it is often found that for photon packets the coherence length as measured by means of the interference pattern extends over almost the whole length of the packet as measured by means of the distance the light travels during the mean life of the decaying state. In these cases the packet length is always close to its lower bound. In the case of wavepackets describing massive particles, however, there is considerable spreading even in the direction of propagation. In the neutron-interference experiment of Kaiser et al.,⁽²⁷⁸⁾ for example, the length $\Delta y(t)$ of the neutron packet at the time of its registration may be larger than the measured coherence length $\Delta_c y = 20 \text{ \AA}$ by more than a factor of 10^5 .

APPENDIX B: PROOF OF THE BELL INEQUALITY

In this appendix we derive the inequality (56)

$$K := |E(a, b) + E(a, b') + E(a', b) - E(a', b')| \leq 2 \quad (\text{B1})$$

of Sec. 5.3 following Bell.⁽¹⁹⁴⁾ $E(a, b)$ is the expectation of the product $r_A r_B$ of two dichotomic variables, r_A and r_B , either of which can take on the values $+1$ and -1 only:

$$E(a, b) := P(+, +|a, b) + P(-, -|a, b) - P(+, -|a, b) - P(-, +|a, b) \quad (\text{B2})$$

and

$$P(r_A, r_B|a, b) = \int P(r_A, r_B|a, b, \lambda) f(\lambda) d\lambda \quad (\text{B3})$$

$$r_A, r_B \in \{-1, +1\} \quad (\text{B4})$$

$$P(r_A, r_B|a, b, \lambda) = P_1(r_A|a, \lambda) P_2(r_B|b, \lambda) \quad (58), (\text{B5})$$

$$f(\lambda) \geq 0, \quad \int f(\lambda) d\lambda = 1. \quad (59), (\text{B6})$$

Since $f(\lambda)$ does not depend on r_A and r_B it is possible to write (B2) as

$$E(a, b) = \int d\lambda f(\lambda) \quad (\text{B7})$$

$$\begin{aligned} & \times \{P_1(+|a, \lambda) P_2(+|b, \lambda) + P_1(-|a, \lambda) P_2(-|b, \lambda) \\ & - P_1(+|a, \lambda) P_2(-|b, \lambda) - P_1(-|a, \lambda) P_2(+|b, \lambda)\}, \end{aligned}$$

and since the first factor on the right-hand side of condition (B5) does not depend on r_B , nor the second on r_A , we may factorize the integrand

$$E(a, b) = \int d\lambda f(\lambda) \{P_1(+|a, \lambda) - P_1(-|a, \lambda)\} \{P_2(+|b, \lambda) - P_2(-|b, \lambda)\} \quad (\text{B8})$$

$$= \int d\lambda f(\lambda) \bar{A}(a, \lambda) \bar{B}(b, \lambda) \quad (\text{B9})$$

where

$$\begin{aligned} \bar{A}(a, \lambda) &= P_1(+|a, \lambda) - P_1(-|a, \lambda) \\ \bar{B}(b, \lambda) &= P_2(+|b, \lambda) - P_2(-|b, \lambda). \end{aligned} \quad (\text{B10})$$

The probability nature of P_1 and P_2 means that

$$0 \leq P_1 \leq 1, \quad 0 \leq P_2 \leq 1, \quad (\text{B11})$$

hence

$$|\bar{A}(a, \lambda)| \leq 1, \quad |\bar{B}(b, \lambda)| \leq 1. \quad (\text{B12})$$

Using (B9) and the fact that $\bar{A}(a, \lambda)$ does not depend on b [nor $\bar{B}(b, \lambda)$ on a] we have

$$E(a, b) \pm E(a, b') = \int d\lambda f(\lambda) \bar{A}(a, \lambda) [\bar{B}(b, \lambda) \pm \bar{B}(b', \lambda)], \quad (\text{B13})$$

and using (B12) for $\bar{A}(a, \lambda)$, (B6) for $f(\lambda)$ and the fact that $f(\lambda)$ does not depend on a and b we obtain

$$|E(a, b) \pm E(a, b')| \leq \int d\lambda f(\lambda) |\bar{B}(b, \lambda) \pm \bar{B}(b', \lambda)|. \quad (\text{B14})$$

Likewise

$$|E(a', b) \mp E(a', b')| \leq \int d\lambda f(\lambda) |\bar{B}(b, \lambda) \mp \bar{B}(b', \lambda)|. \quad (\text{B15})$$

Now, we have

$$|\bar{B}(b, \lambda) \pm \bar{B}(b', \lambda)| + |\bar{B}(b, \lambda) \mp \bar{B}(b', \lambda)| = 2 \max(|\bar{B}(b, \lambda)|, |\bar{B}(b', \lambda)|). \quad (\text{B16})$$

This can be seen by observing that

$$|x \pm y| + |x \mp y| = 2 \max(|x|, |y|), \quad (\text{B17})$$

which in turn may be obtained by considering separately the various possible cases of positive and negative x and y . For example, for x positive and y negative it is $x = |x|, y = -|y|$ and $|x + y| = ||x| - |y|| = \max(|x| - |y|, |y| - |x|)$, and $|x - y| = ||x| + |y|| = |x| + |y|$, so that $|x + y| + |x - y| = \max(|x| - |y| + |x| +$

$|y|, |y| - |x| + |x| + |y| = \max(2|x|, 2|y|)$. Then (B12) means $|x| \leq 1, |y| \leq 1$ so that (B16) with (B12) can be written as

$$|\bar{B}(b, \lambda) \pm \bar{B}(b', \lambda)| + |\bar{B}(b, \lambda) \mp \bar{B}(b', \lambda)| \leq 2.$$

With the normalization (B6) in (B14) and (B15) we arrive at

$$|E(a, b) \pm E(a, b')| + |E(a', b) \mp E(a', b')| \leq 2 \quad (\text{B18})$$

which includes the desired Bell inequality (56), (B1).

APPENDIX C: EPR JOINT PROBABILITY FORMULAS

Here we derive formula (50) of Sec. 5.2

$$P(r_A, r_B | a, b) = \frac{1}{4}(1 - r_A r_B \cos \vartheta) \quad (50)$$

from the rules of quantum mechanics. We have two spin- $\frac{1}{2}$ particles (similar or not) in a state of zero total spin (spin singlet state). The particles move in opposite directions and each enters a Stern-Gerlach-type apparatus where it is deflected upwards or downwards with respect to the axis of its respective apparatus (cf. Fig. 4 in Sec. 5.1). Particle 1 enters apparatus A, which has its axis in the direction of the unit vector \mathbf{a} , and particle 2 enters apparatus B with axis \mathbf{b} . We first calculate the joint probability $P(+, - | a, b)$ that A obtains an up deflection and B a down deflection.

The wave function of the two-particle system in the spin singlet state is (49)

$$\Psi(1, 2) = \frac{1}{\sqrt{2}}|+\rangle^{(1)}|-\rangle^{(2)} - \frac{1}{\sqrt{2}}|-\rangle^{(1)}|+\rangle^{(2)} \quad (\text{C1})$$

where the up $|+\rangle$ and down $|-\rangle$ eigenfunctions of the spin component refer to a fixed but arbitrary axis. Only the spin part needs to be considered here. It happens to be antisymmetric, independent of whether it refers to similar or dissimilar particles. The physical justification for applying formula (C1) in our interpretation is that the two spin- $\frac{1}{2}$ particles are similar and form a coalesced system. If they were dissimilar the conclusions drawn from formula (C1) concerning the nonlocal EPR correlations according to our interpretation would be wrong.

Let the two eigenfunctions of the one-particle spin component operator of apparatus A be $|a+\rangle$ and $|a-\rangle$, and those of apparatus B $|b+\rangle$ and $|b-\rangle$. The operation of a Stern-Gerlach-type apparatus on either of the two particles leads to either $|a+\rangle^{(1)}|b-\rangle^{(2)}$ or $|b-\rangle^{(1)}|a+\rangle^{(2)}$ with equal probability. Hence

$$P(+, - | a, b) = \frac{1}{2}|\langle a+ |^{(1)} \langle b- |^{(2)}, \Psi(1, 2) \rangle|^2 + \frac{1}{2}|\langle b- |^{(1)} \langle a+ |^{(2)}, \Psi(1, 2) \rangle|^2, \quad (\text{C2})$$

whether the particles are similar or not. In the case of dissimilar particles the factors $1/2$ reflect the fact that we are not interested in distinguishing the particles but only in the average result. In the case of similar particles there is now no longer

coalescence and hence no wave function symmetrization, but still the probability expressions should be invariant under particle-label permutation (Sec. 3.1), and this is now the reason for the factors $1/2$.

To be able to evaluate expression (C2) we need the eigenfunctions $|a+\rangle, |b-\rangle$ etc. in terms of the eigenfunctions $|+\rangle, |-\rangle$, which refer to the fixed but arbitrary axis employed in (C1). We let this axis coincide with A's axis \mathbf{a} , so that

$$|a+\rangle = |+\rangle, \quad |a-\rangle = |-\rangle. \quad (\text{C3})$$

Let the axis of B form an angle ϑ with the axis of A. We then have to express the eigenfunctions $|b+\rangle, |b-\rangle$ in the rotated system B in terms of the eigenfunctions of system A^(279,280)

$$\begin{aligned} |b+\rangle &= \cos \frac{\vartheta}{2} \exp[i(\beta + \gamma)/2] |+\rangle + i \sin \frac{\vartheta}{2} \exp[-i(\beta - \gamma)/2] |-\rangle \\ |b-\rangle &= i \sin \frac{\vartheta}{2} \exp[i(\beta - \gamma)/2] |+\rangle + \cos \frac{\vartheta}{2} \exp[-i(\beta + \gamma)/2] |-\rangle. \end{aligned} \quad (\text{C4})$$

The angles β and γ define possible rotations of the other axes but will not appear in the final formulas. Inserting (C1), (C3) and (C4) into (C2) and observing the orthonormality of the functions $|+\rangle, |-\rangle$ for the respective particles we obtain

$$P(+, -|a, b) = \frac{1}{2} \left(\cos \frac{\vartheta}{2} \right)^2 = \frac{1}{4}(1 + \cos \vartheta). \quad (\text{C5})$$

Proceeding in the same way in the other cases (A up, B up; A down, B up; A down, B down) we obtain

$$P(-, +|a, b) = P(+, -|a, b), \quad (\text{C6})$$

$$P(+, +|a, b) = P(-, -|a, b) = \frac{1}{2} \left(\sin \frac{\vartheta}{2} \right)^2 = \frac{1}{4}(1 - \cos \vartheta). \quad (\text{C7})$$

(C5), (C6) and (C7) may be summarized in the form

$$P(r_A, r_B|a, b) = \frac{1}{4}(1 - r_A r_B \cos \vartheta) \quad (\text{C8})$$

where $r_A, r_B \in \{-1, +1\}$. This is formula (50) of Sec. 5.2.

It is amusing to notice that we can also obtain formula (50) by proceeding *as if* the following situation were to hold: after any single interaction either of the two protons of Fig. 4 has a definite direction of spin (i.e. is a spin-up eigenfunction of some $s_{z'}$, cf. Sec. 4.4), say $\vec{\sigma}$ and $-\vec{\sigma}$, respectively, where $\vec{\sigma}$ is a unit vector, and the total spin is zero. The direction of $\vec{\sigma}$ varies from one interaction to the other in such a way that there is spherical symmetry on the average. The Stern-Gerlach-type apparatus which obtains its proton first, say A, turns the spin of its proton into either up or down direction with respect to its axis, say into direction $+\mathbf{a}$, and at

the same time turns the spin of the other proton into the opposite direction $-\mathbf{a}$. Thus here angular momentum is conserved within the system of the two protons, and the apparatuses are not involved in angular-momentum conservation. Then B's apparatus turns the spin of its proton from direction $-\mathbf{a}$ into either up or down direction with respect to B's axis \mathbf{b} , without, however, influencing the spin of proton 1 any more. Here, angular-momentum conservation involves proton 2 and apparatus B, as mentioned in Sec. 4.4 on the Stern-Gerlach experiment.

The spin direction $\vec{\sigma}$ plays the role of the parameter λ in the general consideration of Sec. 5.3. From the point of view of quantum mechanics the parameter $\vec{\sigma}$, interpreted in the above way, is hidden. Of course, in the one-particle states $|+\rangle$ etc. $\vec{\sigma}$ is not hidden but explicitly specifies the spin direction, but the two-particle state $\Psi(1, 2)$ of (C1), which is the only state existing after the interaction at O, has spherical symmetry, and there can be no parameter specifying any direction in such a state in quantum mechanics. This is why we used the proviso "*as if*".

In order to prove our above assertion regarding the derivation of (50) we observe that the conditional probability of obtaining the result r_A (i.e. either $+$ or $-$) with respect to the axis \mathbf{a} , given that the spin of the proton before it enters the apparatus points in the direction \mathbf{b} , is, by (C3) and (C4),

$$P(r_A|\mathbf{a}, \mathbf{b}) = |\langle ar_A|b+\rangle|^2 = |\langle r_A|b+\rangle|^2 = \frac{1}{2}(1 + r_A \mathbf{b}\mathbf{a}),$$

and if we replace the axis \mathbf{b} by the axis $\vec{\sigma}$ and the angle ϑ between \mathbf{a} and \mathbf{b} by the angle α between \mathbf{a} and $\vec{\sigma}$ we get the probability that A obtains the result r_A given that the spin of the proton before it entered the apparatus pointed in the direction $\vec{\sigma}$

$$P(r_A|\mathbf{a}, \vec{\sigma}) = \frac{1}{2}(1 + r_A \vec{\sigma}\mathbf{a}) = \frac{1}{2}(1 + r_A \cos \alpha). \quad (\text{C9})$$

However, the probability of B obtaining the result r_B is not the analogous formula $(1 + r_B(-\vec{\sigma})\mathbf{b})/2$, because A has turned not only the spin of proton 1 into the direction $r_A\mathbf{a}$ but also the spin of proton 2 into direction $-r_A\mathbf{a}$ and then separated the two protons. Thus here we have to replace $-\vec{\sigma}$ by $-r_A\mathbf{a}$, and B's probability is

$$P(r_B|\mathbf{b}, -r_A\mathbf{a}) = \frac{1}{2}(1 + r_B(-r_A\mathbf{a})\mathbf{b}). \quad (\text{C10})$$

The conditional joint probability of A obtaining r_A and B obtaining r_B is given by the product of (C9) with (C10)

$$P(r_A, r_B|\mathbf{a}, \mathbf{b}, \vec{\sigma}) = \frac{1}{2}(1 + r_A \vec{\sigma}\mathbf{a}) \frac{1}{2}(1 - r_A r_B \mathbf{a}\mathbf{b}). \quad (\text{C11})$$

If we integrate over all directions of $\vec{\sigma}$, assuming an isotropic distribution, we obtain

$$\begin{aligned} P(r_A, r_B|\mathbf{a}, \mathbf{b}) &= \frac{1}{4\pi} \int_{-\pi}^{+\pi} d\varphi \int_0^\pi \sin \alpha d\alpha \frac{1}{4} (1 + r_A \cos \alpha) (1 - r_A r_B \cos \vartheta) \\ &= \frac{1}{8} (1 - r_A r_B \cos \vartheta) \int_0^\pi \sin \alpha (1 + r_A \cos \alpha) d\alpha \end{aligned} \quad (\text{C12})$$

where the system of coordinates (x, y, z) for the integration is chosen such that the z axis is in the direction \mathbf{a} , and \mathbf{b} lies in the x - z plane:

$$\mathbf{a} = (0, 0, 1), \quad \mathbf{b} = (\sin \vartheta, 0, \cos \vartheta),$$

so that

$$\vec{\sigma} = (\sin \alpha \cos \varphi, \sin \alpha \sin \varphi, \cos \alpha),$$

$$\mathbf{a}\mathbf{b} = \cos \vartheta, \quad \vec{\sigma}\mathbf{a} = \cos \alpha, \quad \vec{\sigma}\mathbf{b} = \sin \vartheta \cos \varphi \sin \alpha + \cos \vartheta \cos \alpha.$$

It is not difficult to verify that (C12) leads to formula (C8) or (50). Such an “*as-if*” procedure is in fact possible in any EPR situation, not only in that of two spin- $\frac{1}{2}$ particles in the singlet state.

Had we assumed, in a “semiclassical” model, that the two protons were completely independent after their interaction at O, but that the probability formulas for the single protons were still those of standard quantum mechanics, (C9), then proton 1 would arrive at A with $\vec{\sigma}$, proton 2 at B with $-\vec{\sigma}$, and the conditional joint probability would be

$$P_{\text{SC}}(r_A, r_B | \mathbf{a}, \mathbf{b}, \vec{\sigma}) = \frac{1}{2}(1 + r_A \vec{\sigma}\mathbf{a}) \frac{1}{2}(1 - r_B \vec{\sigma}\mathbf{b}). \quad (\text{C13})$$

The integration over an isotropic $\vec{\sigma}$ would then result in

$$\begin{aligned} P_{\text{SC}}(r_A, r_B | \mathbf{a}, \mathbf{b}) &= \frac{1}{4\pi} \int d\varphi \int \sin \alpha d\alpha \frac{1}{2}(1 + r_A \vec{\sigma}\mathbf{a}) \frac{1}{2}(1 - r_B \vec{\sigma}\mathbf{b}) \\ &= \frac{1}{4\pi} \int d\varphi \int \sin \alpha d\alpha \frac{1}{4}(1 + r_A \cos \alpha) (1 - r_B [\sin \vartheta \cos \varphi \sin \alpha + \cos \vartheta \cos \alpha]) \\ &= \frac{1}{4}(1 - \frac{1}{3}r_A r_B \cos \vartheta). \end{aligned} \quad (\text{C14})$$

Expression (C14) differs from the quantum-mechanical formula (C8) only by the factor $1/3$ inside the bracket. This has however the consequence that the expectation (B2) is now

$$E(a, b) = -\frac{1}{3} \cos \vartheta \quad (\text{C15})$$

and the Bell inequality (50)

$$|E(a, b) + E(a, b') + E(a', b) - E(a', b')| \leq 4/3 < 2$$

is always satisfied. We conjecture that such semiclassical formulas would describe the correlations between two dissimilar spin- $\frac{1}{2}$ particles in a state of zero total spin, corresponding in essence to Furry’s⁽²⁸¹⁾ “assumption and method A”, or to Schrödinger’s⁽²⁸²⁾ suggestion of disregarding the phase relations between the eigenfunctions in (C1).

It is seen that the variable $\vec{\sigma}$, which corresponds to λ of Sec. 5.3, in the semiclassical case is independent of \mathbf{a} and \mathbf{b} , but in the quantum-mechanical case is

not, because here the value of $\vec{\sigma}$ for proton 2 before it enters B's apparatus is influenced by A's variable \mathbf{a} , $\vec{\sigma}$ for proton 2 being either $+\mathbf{a}$ or $-\mathbf{a}$.

APPENDIX D: EPR PROBABILITIES IN DIFFERENT SYSTEMS OF EIGENFUNCTIONS

We shall show here that no faster-than-light warning system can be built with devices that obey the formulas of quantum mechanics.^(283–287) The proof uses the fact that in quantum mechanics the apparatuses are represented by operators. Different apparatuses mean different operators, and different operators in general mean different systems of eigenfunctions, and these can be transformed into one another. The normalized quantum-mechanical wave function for a system of two similar particles can be written in the form

$$\Psi_{\text{SA}}(1, 2) = C \sum_{k=1}^{\infty} (\zeta_k(1) u_k(2) \pm \zeta_k(2) u_k(1)), \quad (\text{D1})$$

which is formula (47) from Sec. 5.1 with a properly symmetrized function. The $u_k(x)$ form a complete set of orthonormal eigenfunctions of some operator representing the apparatus of experimenter B. C is a real overall normalization constant, which need not be equal to $1/\sqrt{2}$ because the functions $\zeta_k(x)$, which describe the particle at A, are not presupposed to be orthonormal. With the expansion

$$\zeta_k(x) = \sum_l a_{lk} w_l(x), \quad (\text{D2})$$

where the $w_l(x)$ form a complete set of orthonormal eigenfunctions of some operator representing the apparatus of experimenter A, (D1) can be written as

$$\Psi_{\text{SA}}(1, 2) = C \sum_{lk} a_{lk} (w_l(1)u_k(2) \pm w_l(2)u_k(1)). \quad (\text{D3})$$

We first calculate the probability $P_1(u_n, w_m)$ of a transition where the state (D3) changes into either the state $w_m(1)u_n(2)$ or the state $w_m(2)u_n(1)$ with equal probability. This is the probability that in B's apparatus there will be a particle ("whichever of the two it is") with state u_n and in A's apparatus a particle with state w_m . It is [see the remark on (C2) in Appendix C]

$$\begin{aligned} P_1(u_n, w_m) &= \frac{1}{2} \left| \langle C \sum_{kl} a_{lk} (w_l(1)u_k(2) \pm w_l(2)u_k(1)), w_m(1)u_n(2) \rangle \right|^2 \\ &\quad + \frac{1}{2} \left| \langle C \sum_{kl} a_{lk} (w_l(1)u_k(2) \pm w_l(2)u_k(1)), w_m(2)u_n(1) \rangle \right|^2 \\ &= \frac{C^2}{2} \left| \sum_{kl} a_{lk}^* (\langle w_l(1)u_k(2), w_m(1)u_n(2) \rangle \pm \langle w_l(2)u_k(1), w_m(1)u_n(2) \rangle) \right|^2 \end{aligned}$$

$$+ \frac{C^2}{2} \left| \sum_{kl} a_{lk}^* (\langle w_l(2)u_k(1), w_m(2)u_n(1) \rangle \pm \langle w_l(1)u_k(2), w_m(2)u_n(1) \rangle) \right|^2. \quad (D4)$$

With $\langle w_l(1)u_k(2), w_m(1)u_n(2) \rangle = \langle w_l(2)u_k(1), w_m(2)u_n(1) \rangle = \delta_{lm}\delta_{kn}$ this becomes

$$P_1(u_n, w_m) = \frac{C^2}{2} \left| a_{mn}^* \pm \sum_{kl} a_{lk}^* \langle w_l(2), u_n(2) \rangle \langle u_k(1), w_m(1) \rangle \right|^2 \\ + \frac{C^2}{2} \left| a_{mn}^* \pm \sum_{kl} a_{lk}^* \langle w_l(1), u_n(1) \rangle \langle u_k(2), w_m(2) \rangle \right|^2. \quad (D5)$$

As the scalar products in (D5) are zero the formula reduces to

$$P_1(u_n, w_m) = C^2 |a_{mn}|^2. \quad (D6)$$

The scalar products are zero because the final wave function u_l of the particle in apparatus B and the final wave function w_l of the particle in apparatus A are well separated from each other and do not overlap. The interaction of a wavepacket from the coalesced system (D1) with that apparatus that operated first had led to reduction and to splitting of the system.

Second we consider the probability of B observing that his particle assumes the state u_n irrespective of the state of A's particle. This is obtained by summing the probability $P_1(u_n, w_m)$ over all states of A's particle

$$P_2(u_n) = C^2 \sum_m |a_{mn}|^2, \quad (D7)$$

and we want to show that this probability is unchanged when A uses a different apparatus. Let the eigenfunctions of the new operator corresponding to the new apparatus be $w'_m(x)$. They are related to the eigenfunctions $w_m(x)$ of the original operator by

$$w_m(x) = \sum_k U_{mk} w'_k(x). \quad (D8)$$

U_{mk} is a unitary matrix ($\sum_m U_{jm}^* U_{km} = \delta_{jk}$), and the index k may even be continuous and the sum an integral. Actually, the transformation (D8) need not even be unitary,^(287,288) but we will not pursue this here. By inserting (D8) into (D3) we can write $\Psi_{SA}(1, 2)$ in the form

$$\Psi_{SA}(1, 2) = C \sum_{mn} a_{mn} \left(u_n(2) \sum_k U_{mk} w'_k(1) \pm u_n(1) \sum_k U_{mk} w'_k(2) \right) \\ = C \sum_{lk} \underbrace{\sum_j a_{jk} U_{jl}}_{b_{lk}} (w'_l(1)u_k(2) \pm w'_l(2)u_k(1)). \quad (D9)$$

The probability that this changes into either $w'_m(1)u_n(2)$ or $w'_m(2)u_n(1)$ is $C^2 |b_{mn}|^2$, by analogy with (D3) and (D7). Hence

$$\begin{aligned} P_2(u_n) &= C^2 \sum_m |b_{mn}|^2 = C^2 \sum_m b_{mn}^* b_{mn} = C^2 \sum_{mjk} a_{jn}^* U_{jm}^* a_{kn} U_{km} \\ &= C^2 \sum_{jk} a_{jn}^* a_{kn} \sum_m U_{jm}^* U_{km} = C^2 \sum_k |a_{kn}|^2, \end{aligned} \quad (\text{D10})$$

and this coincides with (D7), concluding the proof of our assertion.

Finally we want to show that even if A chooses not to use his apparatus and to do nothing this will make no difference. In this case, when B's particle assumes the state u_n , A's particle will assume some correlated state ζ_n . The probability of B's particle assuming the state u_n is the probability of the transition where the state $\Psi_{\text{SA}}(1, 2)$ changes either into the state $\zeta_n(1)u_n(2)$ or into the state $\zeta_n(2)u_n(1)$. When ζ_n is expressed in terms of the w_m , according to formula (D2), the two states become

$$\sum_m a_{mn} w_m(1) u_n(2) \quad \text{and} \quad \sum_m a_{mn} w_m(2) u_n(1),$$

respectively, and the transition probability becomes

$$\begin{aligned} P'_2(u_n) &= \frac{1}{2} \left| \langle \Psi_{\text{SA}}(1, 2), \sum_m a_{mn} w_m(1) u_n(2) \rangle \right|^2 \times \left| \sum_m a_{mn} w_m(1) u_n(2) \right|^{-2} \\ &+ \frac{1}{2} \left| \langle \Psi_{\text{SA}}(1, 2), \sum_m a_{mn} w_m(2) u_n(1) \rangle \right|^2 \times \left| \sum_m a_{mn} w_m(2) u_n(1) \right|^{-2}. \end{aligned} \quad (\text{D11})$$

The denominators are different from 1 because the a_{mn} come in via the $\zeta_n(x)$ in formula (D2), and the ζ 's are not normalized. With (D3) the first term of expression (D11) becomes

$$\begin{aligned} T_1 &= \frac{C^2}{2} \\ &\times \left| \sum_{klm} a_{lk}^* a_{mn} (\langle w_l(1), w_m(1) \rangle \langle u_k(2), u_n(2) \rangle \pm \langle w_l(2), u_n(2) \rangle \langle u_k(1), w_m(1) \rangle) \right|^2 \\ &\times \left| \sum_{mk} a_{mn}^* a_{kn} \langle w_m(1) u_n(2), w_k(1) u_n(2) \rangle \right|^{-1} \\ &= \frac{C^2}{2} \left| \sum_m a_{mn}^* a_{mn} \right|^2 \times \left| \sum_m a_{mn}^* a_{mn} \right|^{-1} = \frac{C^2}{2} \sum_m |a_{mn}|^2. \end{aligned} \quad (\text{D12})$$

The second term of (D11) leads to the same expression, so

$$P'_2(u_n) = C^2 \sum_m |a_{mn}|^2 = P_2(u_n), \quad (\text{D13})$$

which is what we wanted to show.

Received 25 January 1994

Résumé

On présente une alternative à l'interprétation de Copenhague du formalisme de la mécanique quantique non-relativiste. La différence fondamentale est que l'interprétation nouvelle est formulée dans le langage du réalisme épistémologique, ce qui inclue le changement de certains concepts fondamentaux de la physique. La fonction ψ n'est plus interprétée comme une amplitude de probabilité qui se réfère au comportement observé d'une particule élémentaire, mais comme un champ physique objectif représentant la particule elle-même. Ainsi, les particules sont des objets étendus et leur extension varie avec le temps, conformément à la variation de ψ . Elles sont considérées comme des régions fondamentales de l'espace, sans structure interne. Ceci implique un certain type de non-localité. La symétrisation de la fonction d'onde dans l'espace de configuration est interprétée comme une description mathématique d'un processus physique, qui conduit aussi aux effets non-locaux. Une attention particulière est donnée au problème de mesure, au processus de réduction, au chat de Schrödinger, à l'ami de Wigner, aux corrélations de Einstein-Podolsky-Rosen, à la quantification du champ et aux distributions de la statistique quantique. Des expériences permettant de différencier l'interprétation proposée ici avec celle de Copenhague sont présentées.

References

1. M. Jammer, *The Conceptual Development of Quantum Mechanics* (McGraw-Hill, New York, 1966).
2. *Ibid.*, Chap. 7.
3. W. Heisenberg, Z. Phys. **43**, 172 (1927). The quotation is from p. 185.
4. J.C. Maxwell, *A Treatise on Electricity and Magnetism* (Clarendon Press, Oxford, 1891), Vol 2, Sec. 528. Republished by Dover, New York, 1954, p. 176.
5. H. Epstein and V. Glaser, Ann. Inst. Poincaré A **19**, 211 (1973); G. Scharf, *Finite Quantum Electrodynamics*, 2nd ed. (Springer-Verlag, Berlin, 1995).
6. J.M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons*, 2nd ed. (Springer-Verlag, New York, 1976), Sec. 16-5.
7. J.D. Bjorken and S.D. Drell, *Relativistic Quantum Fields* (McGraw-Hill, New York, 1965).
8. *Ibid.*, Sec. 11.1.
9. P.A.M. Dirac, Fields and Quanta **3**, 139 (1972). The quotation is from pp. 163, 164; see also P.A.M. Dirac in *Perturbative Quantum Chromodynamics*, edited by D.W. Duke and J.F. Owens (American Institute of Physics, NY, 1981), p. 129.
10. A. Einstein, Dialectica **2**, 320 (1948). Reprinted in Ref. 11, letter of 5 April 1948.
11. M. Born, editor, *Albert Einstein, Hedwig und Max Born, Briefwechsel 1916-1955* (Nymphenburger Verlagsbuchhandlung, München, 1969). English translation: *The Born-Einstein Letters* (MacMillan, London, 1971).
12. P.A. Schilpp, editor, *Albert Einstein: Philosopher - Scientist* (Cambridge University Press, London, 1949).
13. A. Einstein, in Ref. 12, p. 667.
14. A. George, editor, *Louis de Broglie, Physicien et Penseur (Les savants et le monde)* (Albin Michel, Paris, 1953).
15. A. Einstein, in Ref. 14, p. 4. The quotation is from p. 6. My translation.
16. A. Einstein, in *Scientific Papers presented to Max Born on his retirement from the Tait Chair of Natural Philosophy in the University of Edinburgh* (Oliver and Boyd, Edinburgh, London, 1953), p. 33.
17. L. Wittgenstein, *Philosophische Untersuchungen* (Basil Blackwell, Oxford, 1953), Sec. 402. The book contains also the english translation.
18. M. v. Laue, Naturwissenschaften **22**, 439 (1934).
19. E. Schrödinger, *ibid.* **23**, 807, 823, 844 (1935). English translation in: Proc. Am. Philos. Soc. **124**, 323 (1980) and in *Quantum Theory and Measurement*, edited by J.A. Wheeler and W.H. Zurek (Princeton University Press, Princeton, 1983), p. 152.
20. *Idem*, Endeavour **9**, 107 (1950).
21. *Idem*, Sci. Amer. **189**, (3) 52 (1953).
22. M. Planck, *Die Physik im Kampf um die Weltanschauung* (J.A. Barth, Leipzig, 1937). Partly reprinted in English in Ref. 23.
23. L. Jánossy, Acta Phys. Hung. **1**, 423 (1952).

24. M. Jammer, *The Philosophy of Quantum Mechanics* (Wiley, New York, 1974).
25. L. de Broglie, *The Current Interpretation of Wave Mechanics, A Critical Study* (Elsevier, Amsterdam, 1964), p. 22, 37. English translation of *Étude Critique des Bases de l'Interprétation Actuelle de la Mécanique Ondulatoire* (Gauthier-Villars, Paris, 1963).
26. M. Bunge, in Ref. 27, p. 1; M. Bunge and A.J. Kálnay, in *Philosophie und Physik*, edited by R. Haller and J. Götschl (Vieweg, Braunschweig, 1975), p. 25.
27. M. Bunge, editor, *Quantum Theory and Reality* (Studies in the Foundations, Methodology and Philosophy of Science, Vol. 2) (Springer-Verlag, Berlin, 1967).
28. *Idem*, *Foundations of Physics* (Springer-Verlag, New York, 1967).
29. J. Mehra, editor, *The Physicist's Conception of Nature* (Reidel, Boston, 1973).
30. J.S. Bell, *Speakable and unspeakable in quantum mechanics* (Cambridge University Press, Cambridge, 1987).
31. *Idem*, in Ref. 29, p. 687. Reprinted in Ref. 30, p. 40.
32. *Idem*, in *Sixty-Two Years of Uncertainty*, edited by A.I. Miller (Plenum Press, New York, 1990), p. 17. Reprinted in *Physics World* **3**, (8) 33 (1990).
33. J. Rayski, *Found. Phys.* **3**, 89 (1973).
34. J.-M. Lévy-Leblond, *Dialectica* (Neuchâtel) **30**, 161 (1976).
35. P.H. Stapp, *Nuovo Cimento B* **40**, 191 (1977); *Am. J. Phys.* **53**, 306 (1985).
36. H.V. Roberts, *Proc. R. Soc. London A* **360**, 135 (1978).
37. N. Maxwell, *Found. Phys.* **12**, 607 (1982).
38. M.E. Burgos, *Found Phys.* **14**, 739, 753 (1984).
39. R.K. Popper, in Ref. 27, p. 7; in *Open Questions in Quantum Physics*, edited by G. Tarozzi and A. van der Merwe (Reidel, Dordrecht, 1985), p. 3.
40. P. Pearle, *Annals of the New York Academy of Sciences* **480**, 539 (1986).
41. D. Bohm, B.J. Hiley, and P.N. Kaloyerou, *Phys. Reports* **144**, 321 (1987).
42. F. Rohrlich, *Annals of the New York Academy of Sciences* **480**, 373 (1987).
43. J. Dorling, in *Schrödinger: Centenary Celebration of a Polymath*, edited by C.W. Kilmister (Cambridge University Press, Cambridge, 1987), p. 16.
44. D. Dieks, *Conceptus* **22** (57) 31 (1988).
45. *Idem*, *Phys. Rev. A* **49**, 2290 (1994).
46. *Idem*, *Phys. Lett. A* **197**, 367 (1995).
47. A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1972, 1973), Volumes I and II. English translation of *Mécanique Quantique* (Dunod, Paris, 1964), Vols. 1 and 2.
48. D. Bohm, *Phys. Rev.* **85**, 166, 180 (1952).
49. F.J. Dyson, *Sci. Am.* **199** (3) 74 (1958); A. Pais, *Niels Bohr's Times* (Clarendon Press, Oxford, 1991), p. 29.
50. C.G. Wohl, R.N. Cahn, A. Rittenberg, T.G. Trippe, G.P. Yost, F.C. Porter, J.J. Hernandez, L. Montanet, R.E. Hendrick, R.L. Crawford, M. Roos, N.A. Törnqvist, G. Höhler, M. Aguilar-Benitez, T. Shimada, M.J. Losty, G.P. Gopal, Ch. Walck, R.E. Shrock, R. Frosch, L.D. Roper, W.P. Trower, and B. Arm-

- strong, Rev. Mod. Phys. **56**, (2) Part II, S1 (1984) (*Review of particle properties*).
51. W. Heisenberg, Daedalus **87**, 95 (1958).
52. L.E. Ballentine, Rev. Mod. Phys. **42**, 358 (1970).
53. A. Einstein, Philosophy of Science **1**, 163 (1934). The quotation is from p. 168.
54. E. Schrödinger, Naturwissenschaften **14**, 664 (1926).
55. *Idem*, Ann. Phys. (Leipzig) **79**, 489 (especially p. 526); **79**, 734 (especially pp. 754, 755); **81**, 109 (especially Sec. 7) (1926). English translation: *Collected Papers on Wave Mechanics* (Chelsea Publ. Comp., New York, 1982), especially pp. 39, 60, 120.
56. *Idem*, Sitzungsberichte der Preussischen Akademie der Wissenschaften, physikalisch-mathematische Klasse, Berlin, 1930, p. 296.
57. *Idem*, Brit. J. Phil. Sci. **3**, 109, 233 (1952).
58. Ref. 14, p. 16.
59. M. Bunge, Am. J. Phys. **24**, 272 (1956), especially p. 277.
60. S. Brandt and H.D. Dahmen, *The Picture Book of Quantum Mechanics* (Wiley, New York, 1985), Sec. 6.2.
61. D. Bhaumik, B. Dutta-Roy, and G. Ghosh, J. Phys. A: Math. Gen. **19**, 1355 (1986).
62. G. Alber and P. Zoller, Physikalische Blätter **45**, 477 (1989).
63. A. Einstein, in *James Clerk Maxwell: A Commemoration Volume* (Cambridge University Press, Cambridge, 1931), p. 66. Reprinted in *The World as I See It* (Covici-Friede, New York, 1934). German version in *Mein Weltbild*, edited by C. Seelig (Ullstein, Frankfurt, 1955), p. 159.
64. *Idem*, *Über die spezielle und die allgemeine Relativitätstheorie* (Vieweg, Braunschweig, 1956), Anhang 5. English translation: *Relativity, The Special and The General Theory* (Methuen, London, 1970), Appendix V.
65. Ref. 47, formula (XVII.22).
66. *Ibid.*, p. 725.
67. *Ibid.*, Sec. VIII.16.
68. *Ibid.*, Secs. VII.16 to 20, VIII.15 to 19.
69. *Ibid.*, pp. 147, 148.
70. T. Marshall and E. Santos, Found. Phys. **18**, 185 (1988).
71. G.T. Reynolds, K. Spartalian, and D.B. Scarl, Nuovo Cimento B **61**, 355 (1969).
72. L. Mandel, in *Progress in Optics XIII*, edited by E. Wolf (North-Holland, Amsterdam, 1976), p. 27.
73. B.R. Wheaton, *The Tiger and the Shark, Empirical Roots of Wave-Particle Dualism* (Cambridge University Press, Cambridge, 1983).
74. P.A.M. Dirac, Proc. R. Soc. London A **167**, 148 (1938). The quotation is from p. 160.
75. J.P. Vigiér, Lett. Nuovo Cimento **24**, 258 (1979).
76. N. Cufaro Petroni and J.P. Vigiér, Lett. Nuovo Cimento **25**, 151; **26**, 149 (1979).

77. U. Bonse and H. Rauch, editors, *Neutron Interferometry* (Clarendon Press, Oxford, 1979).
78. R. Hofstadter, editor, *Nuclear and Nucleon Structure: a collection of reprints with an introduction* (Benjamin, New York, 1963).
79. R.P. Feynman, *Phys. Today* **19** (8), 31 (1966).
80. F. Rohrlich, in Ref. 29, p. 331.
81. H.A. Lorentz, *The Theory of Electrons and its Applications to the Phenomena of Light and Radiant Heat* (Teubner, Leipzig, 1909 and Dover, New York, 1952), p. 215.
82. E. Schrödinger, *Ann. Phys. (Leipzig)* **82**, 265 (1927).
83. A.O. Barut, *Physica Scripta T* **21**, 18 (1988), and references cited therein.
84. M. Jammer, Ref. 1, p. 254. Quoted from M. Jammer, *Concepts of Mass in Classical and Modern Physics* (Harvard University Press, Cambridge, Massachusetts, 1961), p. 153.
85. L. Landau and R. Peierls, *Z. Phys.* **62**, 188 (1930).
86. R.H. Good, Jr. and T.J. Nelson, *Classical Theory of Electric and Magnetic Fields* (Academic Press, New York, 1971), p. 614.
87. W. Pauli, *Z. Phys.* **80**, 573 (1933).
88. W. Pauli and V.F. Weisskopf, *Helv. Phys. Acta* **7**, 709 (1934).
89. P.A.M. Dirac, *Proc. R. Soc. London A* **180**, 1 (1942).
90. W. Heisenberg, *Z. Naturforsch. A* **9**, 292 (1954).
91. *Idem*, *Introduction to the Unified Field Theory of Elementary Particles* (Wiley, New York, 1966).
92. *Idem*, in *Aspects of Quantum Theory*, edited by A. Salam and E.P. Wigner (Cambridge University Press, Cambridge, 1972), p. 129.
93. N.N. Bogoliubov and E.V. Shirkov, *Introduction to the Theory of Quantized Fields* (Wiley, New York, 1980). English translation of *Vvedenie v teoriyu kvantovannykh polei* (Izdatel'stva Nauka, Moskva, 1976), Chapter 1; and other books on quantum field theory.
94. D. Rapp, *Quantum Mechanics* (Holt, Rinehart and Winston, New York, 1971), p. 20.
95. M. Planck, *The Universe in the Light of Modern Physics* (Allen and Unwin, London, 1934). Reprinted in *The New Science* (Greenwich editions published by Meridian Books Inc., Silver Spring, Maryland, USA, 1959), p. 155, especially p. 188.
96. D. Bohm, Ref. 97, p. 99.
97. D. Bohm, *Quantum Theory* (Prentice-Hall, Englewood Cliffs, New Jersey, 1951).
98. Ref. 47, p. 142, 146.
99. N. Bohr, in Ref. 13, p. 199. The quotation is from p. 233.
100. B.O. Koopman, in *Proc. of Symposia in Applied Math.* (McGraw-Hill, New York, 1957), Vol. VII, *Applied Probability*, p. 97.
101. H.P. Robertson, *Phys. Rev.* **34**, 163 (1929).

102. W. Pauli, in *Handbuch der Physik*, edited by H. Geiger and K. Scheel (Springer-Verlag, Berlin, 1933), Vol. 24, Part I. English version: *General Principles of Quantum Mechanics*, translated by P. Achuthan and K. Venkatesan (Springer-Verlag, Berlin, 1980), Section 14.
103. Ref. 47, Sec. XIV.8.
104. R. Mirman, *Nuovo Cimento B* **18**, 110 (1973).
105. R.G. Parr and W. Yang, *Density-functional theory of atoms and molecules* (Oxford University Press, New York, 1989).
106. Ref. 47, Sec. VII.6.
107. V. Fock, *Z. Phys.* **75**, 622 (1932), formula (5).
108. *Idem*, *Z. Phys.* **75**, 622 (1932); **76**, 852(E) (1932).
109. D.I. Blochinzew, *Grundlagen der Quantenmechanik* (Harry Deutsch, Frankfurt, 1962), Secs. 114 to 116. English version: *Quantum Mechanics* (Reidel, Dordrecht, 1964).
110. S.S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper and Row, New York, 1966), Chapter 6.
111. L.I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1968), Section 54. The statement related to footnote 1 on p. 504 of this book is true only for interactions without retardation.
112. P. Ehrenfest, *Z. Phys.* **78**, 555 (1932).
113. R. Hanbury Brown and R. Twiss, *Nature* **177**, 27 (1956); *Proc. R. Soc. London A* **243**, 291 (1958).
114. E.M. Purcell, *Nature* **178**, 1449 (1956).
115. J. Brendel, E. Mohler, and W. Martienssen, *Phys. Rev. Lett.* **9**, 1142 (1989).
116. P.G. Kwiat, A.M. Steinberg, and R.Y. Chiao, *Phys. Rev. A* **45**, 7729 (1992).
117. T.S. Larchuk, R.A. Campos, J.G. Rarity, P.R. Tapster, E. Jakeman, B.E.A. Saleh, and M.C. Teich, *Phys. Rev. Lett.* **70**, 1603 (1993).
118. X.Y. Zou, T. Grayson, G.A. Barbosa, and L. Mandel, *Phys. Rev. A* **47**, 2293 (1993).
119. J.M. Blatt, *Theory of Superconductivity* (Academic Press, New York, 1964), p. 63.
120. V. Ernst, *Z. Phys.* **229**, 432 (1969).
121. V. Ernst and P. Stehle, *Phys. Rev.* **176**, 1456 (1968).
122. G. Ludwig, in *Werner Heisenberg und die Physik unserer Zeit*, edited by F. Bopp (Vieweg, Braunschweig, 1961), p. 150; in Ref. 29, p. 702.
123. G. Ludwig, *Physikalische Blätter* **11**, 489 (1955).
124. A. Einstein, in *Albert Einstein / Arnold Sommerfeld. Briefwechsel*, edited by A. Hermann (Schwabe & Co., Basel, 1968), letter [49] of 21 August 1926. My translation.
125. A.O. Barut, *Found. Phys. Lett.* **1**, 47 (1988).
126. *Idem*, *Ann. Phys. (Leipzig)* **45**, 31 (1988), and references cited therein.
127. J.D. Bjorken and S.D. Drell, Ref. 7, Section 15.2.
128. D.W. Keith, M.L. Schattenburg, H.I. Smith, and D.E. Pritchard, *Phys. Rev. Lett.* **61**, 1580 (1989).

129. O. Carnal and J. Mlynek, Phys. Rev. Lett. **66**, 2689 (1991).
130. A. Jabs, Brit. J. Phil. Sci. **54**, 405 (1992). In this publication the coalesced system of the present exposition was called a coalesced wavepacket, with internal structurelessness. This is no longer upheld because of the possible existence of dynamic interactions and of the identification of wavepackets with degrees of freedom.
131. *Idem*, *Principles of an Objective Quantum Theory of Wave Packets* (preprint SLAC-PUB-2179, Stanford, USA, 1978).
132. J. v. Neumann, *Mathematische Grundlagen der Quantenmechanik* (Springer-Verlag, Berlin, 1932). English translation: *Mathematical Foundations of Quantum Mechanics* (Princeton University Press, Princeton, 1955).
133. Ref. 47, p. 298.
134. K.S. Tausk, 1966, private communication.
135. V.M. Gusev, O.N. Kompanets, A.R. Kukudzhinov, V.S. Letokhov, and E.L. Mikhailov, Sov. J. Quantum Electron. **4**, 1370 (1975). [Kvant. Elektron. **1**, 2465 (1974)].
136. C. Jönsson, Z. Phys. **161**, 454 (1961). English translation: Am. J. Phys. **42**, 4 (1974).
137. Ref. 47, p. 842.
138. E. Wigner, in Ref. 29, p. 320.
139. D. Bohm, Ref. 97, p. 412.
140. M. Jammer, Ref. 1, p. 324.
141. E. Schrödinger, Ref. 57, p. 113.
142. W. Gerlach and O. Stern, Z. Phys. **9**, 349 (1922).
143. *Idem*, Ann. Phys. (Leipzig) **74**, 673 (1924).
144. *Ibid.*, **74**, 761 (1924).
145. E. Wrede, Z. Phys. **41**, 569 (1927).
146. Ref. 47, Sec. III.10.
147. W. Pauli, in Ref. 148, letter [236] of 4 September 1929 to Knauer.
148. A. Hermann, K. v. Meyenn, and V.F. Weisskopf, editors, *Wolfgang Pauli, Scientific Correspondence with Bohr, Einstein, Heisenberg a.o.* (Springer-Verlag, New York, 1979), Volume I: 1919-1929.
149. F. Knauer, Z. Phys. **59**, 807 (1930).
150. N.F. Mott and H.S.W. Massey, *The Theory of Atomic Collisions* (Clarendon Press, Oxford, 1965).
151. *Ibid.*, p. 214.
152. *Ibid.*, Chap. IX.
153. M. Laméhi-Rachti and W. Mittig, Phys. Rev. D **14**, 2543 (1976).
154. E. Merzbacher, *Quantum Mechanics* (Wiley, New York, 1970), p. 290.
155. Ref. 96, p. 593.
156. R.A. Beth, Phys. Rev. **50**, 115 (1936).
157. P.A.M. Dirac, Ref. 158, p. 12.
158. P.A.M. Dirac, *The Principles of Quantum Mechanics* (Clarendon Press, Oxford, 1967).

- 159. E. Schrödinger, Ref. 19, p. 812.
- 160. E.P. Wigner, in *The Scientist Speculates*, edited by I.J. Good (Heinemann, London, 1961). Reprinted in *Symmetries and Reflections* (Indiana University Press, London, 1967), p. 171.
- 161. H. Margenau and R.N. Hill, Progr. Theor. Phys. **26**, 722 (1961). The quotation is from p. 723.
- 162. J. v. Neumann, Ref. 132, Section III.3.
- 163. A.H. Compton and A.W. Simon, Phys. Rev. **26**, 289 (1925).
- 164. N. Bohr, Naturwissenschaften **17**, 483 (1929).
- 165. J. v. Neumann, Ref. 132, footnote 207.
- 166. M. Jammer, Ref. 1, p. 370.
- 167. A. Einstein, Phys. Z. **18**, 121 (1917).
- 168. P.J. Bussey, Phys. Lett. A **106**, 407 (1984); **118**, 377 (1986); **120**, 51 (1987).
- 169. D. Home and M.A.B. Whitaker, Phys. Lett. A **117**, 439 (1986).
- 170. G. Badurek, H. Rauch, and D. Tuppinger, Phys. Rev. A **34**, 2600 (1986).
- 171. J. Summhammer, H. Rauch, and D. Tuppinger, Phys. Rev. A **36**, 4447 (1987).
- 172. A. Einstein, in Ref. 11, letter of 29 April 1924.
- 173. *Ibid.*, letter of 4 December 1926. My translation.
- 174. *Idem*, Journal of the Franklin Institute **221**, 313 (1936). English translation *ibid.* 349.
- 175. *Idem*, in Ref. 11, letter of 7 September 1944.
- 176. P. Speziali, editor, *Albert Einstein, Michele Besso, Correspondance 1903 - 1955* (Hermann, Paris, 1972).
- 177. *Ibid.*, letter 165 of 30 November 1949 and many earlier letters.
- 178. A. Einstein, in Ref. 176, letter 172 of 15 April 1950.
- 179. W. Pauli, in Ref. 11, letter of 31 March 1954 to Born. My translation.
- 180. J. v. Neumann, Ref. 132, Sections IV.1 and IV.2.
- 181. J.S. Bell, Physics **1**, 195 (1965). Reprinted in Ref. 30, p. 14.
- 182. *Idem*, Rev. Mod. Phys. **38**, 447 (1966). Reprinted in Ref. 30, p. 1.
- 183. P.A.M. Dirac, in Ref. 29, p. 1. The quotation is from p. 7.
- 184. E.T. Jaynes, American Scientist (Burlington, Vermont) **53**, 136 A (1965).
- 185. A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. **47**, 777 (1935).
- 186. D. Bohm, Ref. 97, p. 614.
- 187. Ref. 47, p. 562.
- 188. A. Einstein, in Ref. 12, p. 84, 682, 683.
- 189. N. Herbert, Found. Phys. **12**, 1171 (1982).
- 190. P.W. Milonni and M.L. Hardies, Phys. Lett. A **92**, 321 (1982).
- 191. W.K. Wootters and W.H. Zurek, Nature **299**, 802 (1982).
- 192. D. Dieks, Phys. Lett. A **92**, 271 (1982).
- 193. G.C. Ghirardi and T. Weber, Nuovo Cimento B **78**, 9 (1983).
- 194. J.S. Bell, J. Phys. (Paris) **42**, C2, 41 (1981). Reprinted in: Speculations in Science and Technology **10**, 269 (1987) and also in Ref. 30, p. 139.
- 195. J.F. Clauser and M.A. Horne, Phys. Rev. D **10**, 526 (1974).

196. J.S. Bell, Epistemological Letters (Association Ferdinand Gonseth, Case Postale 1081, CH-2501 Bienne, 1977), Written Symposium, 15th Issue, p. 79. Reprinted in Ref. 30, p. 52.
197. P.H. Stapp, Nuovo Cimento B **40**, 191 (1977). Especially Section 4.5.a.
198. J.F. Clauser and A. Shimony, Rep. Progr. Phys. **41**, 1881 (1978).
199. A.J. Duncan, in *Coherence in Atomic Collision Physics*, edited by H.J. Beyer, K. Blum, and R. Hippler (Pergamon, New York, 1988), p. 321.
200. Z.Y. Ou and L. Mandel, Phys. Rev. Lett. **61**, 50 (1988).
201. Y.H. Shih and C.O. Alley, Phys. Rev. Lett. **61**, 2921 (1988).
202. J.G. Rarity and P.R. Tapster, Phys. Rev. Lett. **64**, 2495 (1990).
203. H. Paul, Fortschr. Phys. **28**, 633 (1980).
204. B.I. Spasskii, A.V. Moskovskii, Sov. Phys. Usp. **27**, 273 (1984) [Usp. Fiz. Nauk **142**, 599 (1984)].
205. A. Aspect, J. Dalibard, and G. Roger, Phys. Rev. Lett. **49**, 1804 (1982).
206. A. Aspect, P. Grangier, and G. Roger, Phys. Rev. Lett. **47**, 460 (1981).
207. *Idem*, Phys. Rev. Lett. **49**, 91 (1982).
208. G. Faraci, D. Gutkowski, S. Notarrigo, and A.R. Pennisi, Lett. Nuovo Cimento **9**, 607 (1974).
209. A.R. Wilson, J. Lowe, and D.K. Butt, J. Phys. G: Nucl. Phys. **2**, 613 (1976).
210. M. Bruno, M. d'Agostino, and C. Maroni, Nuovo Cimento B **40**, 143 (1977).
211. B.S. Cirel'son, Lett. Math. Phys. **4**, 93 (1980).
212. L.J. Landau, Phys. Lett. A **120**, 54 (1987).
213. J.D. Bjorken and S.D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964), Section 7.9.
214. J.M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons*, 2nd ed. (Springer-Verlag, New York, 1976), Sec. 12-2.
215. P. Marmier and E. Sheldon, *Physics of Nuclei and Particles* (Academic Press, New York, 1970), Vol. II, p. 902.
216. H. Arenhövel and M. Sanzone, *Photodisintegration of the Deuteron: A Review of Theory and Experiment* (Few-Body Systems, Suppl. 3), (Springer-Verlag, Wien, 1991), Secs. 6 and 7.
217. K.M. Schmitt, P. Wilhelm, and H. Arenhövel, Few-Body Systems **10**, 105 (1991), Figs. 1, 12, 14.
218. M.L. Rustgi and L.J. Pandey, Phys. Rev. C **40**, 1581 (1989).
219. A.O. Barut and M. Božić, Nuovo Cimento B **101**, 595 (1988).
220. J.M. Jauch, in *Foundations of Quantum Mechanics*, edited by B. d'Espagnat (Academic Press, New York, 1971) (Proc. Intern. School of Physics "Enrico Fermi", Course 49, Varenna, Italy, 1970), p. 20.
221. W. Heisenberg, Naturwissenschaften **14**, 989 (1926).
222. *Idem*, in Ref. 148, letter no. [142] of 28 July 1926.
223. P.A.M. Dirac, Proc. R. Soc. London A **114**, 243 (1927).
224. M. Born and P. Jordan, Z. Phys. **34**, 858 (1925).
225. M. Born, W. Heisenberg, and P. Jordan, Z. Phys. **35**, 557 (1926).
226. P. Jordan, Z. Phys. **44**, 473 (1927).

- 227. *Ibid.* **45**, 766 (1927).
- 228. *Ibid.* **75**, 648 (1932).
- 229. P. Jordan and O. Klein, Z. Phys. **45**, 751 (1927).
- 230. P. Jordan and E. Wigner, Z. Phys. **47**, 631 (1928).
- 231. W. Heisenberg and W. Pauli, Z. Phys. **56**, 1 (1929).
- 232. *Ibid.* **59**, 168 (1930).
- 233. M. Planck, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin (meeting of 18 May 1899), p. 440. Reprinted in *Physikalische Abhandlungen und Vorträge* (Vieweg, Braunschweig, 1958), Vol. I, p. 560.
- 234. Lord Rayleigh, Philos. Mag. **49**, 539 (1900); Nature **72**, 54 (1905).
- 235. J.H. Jeans, Philos. Mag. **10**, 91 (1905).
- 236. M. v. Laue, Ann. Phys. (Leipzig) **44**, 1197 (1914).
- 237. S.N. Bose, Z. Phys. **26**, 178 (1924).
- 238. M. Planck, *Vorlesungen über die Theorie der Wärmestrahlung* (J.A. Barth, Leipzig, 1906), 1st edition, p. 154, 178.
- 239. A. Landé, Z. Phys. **33**, 571 (1925).
- 240. P. Ehrenfest, Phys. Z. **7**, 528 (1906).
- 241. P. Debye, Ann. Phys. (Leipzig) **33**, 1427 (1910).
- 242. E. Schrödinger, Phys. Z. **25**, 89 (1924).
- 243. L. Mandel, J. Opt. Soc. Am. **51**, 797 (1961).
- 244. A. Joffé, Ann. Phys. (Leipzig) **36**, 534 (1911).
- 245. J. Ishiwara, Phys. Z. **13**, 1142 (1912).
- 246. M. Wolfke, Phys. Z. **22**, 375 (1921).
- 247. L. de Broglie, Comptes Rendus Acad. Sci. (Paris) **175**, 811 (1922).
- 248. W. Bothe, Z. Phys. **20**, 145 (1923); **23**, 214 (1924).
- 249. *Ibid.* **41**, 345 (1927).
- 250. A. Einstein, Sitzungsberichte der Preussischen Akademie der Wissenschaften, physikalisch-mathematische Klasse, Berlin, 1925, p. 3.
- 251. E. Fermi, Z. Phys. **36**, 902 (1926).
- 252. P. Jordan, Z. Phys. **41**, 711 (1927).
- 253. L.S. Ornstein and H.A. Kramers, Z. Phys. **42**, 481 (1927).
- 254. L. Oster, Am. J. Phys. **38**, 754 (1970).
- 255. S.K. Ma, *Statistical Mechanics* (World Scientific, Singapore, 1985), Sec. 3.2.
- 256. R.N. Zitter and R.C. Hilborn, Am. J. Phys. **55**, 522 (1987).
- 257. F.S. Crawford, Am. J. Phys. **56**, 883 (1988).
- 258. W. Pauli, Z. Phys. **18**, 272 (1923).
- 259. A. Einstein and P. Ehrenfest, Z. Phys. **19**, 301 (1923).
- 260. P.A.M. Dirac, Proc. R. Soc. London A **106**, 581 (1924).
- 261. L. Natanson, Phys. Z. **12**, 659 (1911).
- 262. A. Einstein, Ann. Phys. (Leipzig) **17**, 132 (1905).
- 263. L. Mandel, Proc. Phys. Soc. London **74**, 233 (1959).
- 264. W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1970), Vol. 1.

- 265. *Ibid.*, pp. 173, 268.
- 266. *Ibid.*, Chapter VI.
- 267. B.R. Mollow, Phys. Rev. **168**, 1896 (1968), formula (3.24b).
- 268. Ref. 264, p. 266.
- 269. R. Loudon, *The quantum theory of light* (Oxford University Press, Oxford, 1983), Secs. 6.7 and 6.8.
- 270. F.T. Arecchi, Phys. Rev. Lett. **15**, 912 (1965).
- 271. M. Bauer and P.A. Mello, Ann. Phys. (New York) **111**, 38 (1978).
- 272. J. Hilgevoord and J.B.M. Uffink, Phys. Lett. A **95**, 474 (1983).
- 273. H.M. Bradford, Am. J. Phys. **44**, 1058 (1976).
- 274. C. Almeida and A. Jabs, Am. J. Phys. **52**, 921 (1984).
- 275. J.B.M. Uffink and J. Hilgevoord, Phys. Lett. A **105**, 176 (1984); J.B.M. Uffink, *ibid.* **108**, 59 (1985).
- 276. A.G. Klein, G.I. Opat, and W.A. Hamilton, Phys. Rev. Lett. **50**, 563 (1983).
- 277. A. Jabs and R. Ramos, Phys. Rev. Lett. **58**, 2274 (1987).
- 278. H. Kaiser, S.A. Werner, and E.A. George, Phys. Rev. Lett. **50**, 560 (1983).
- 279. R.P. Feynman, R.B. Leighton, and M. Sands, *The Feynman Lectures on Physics* (Addison-Wesley, Reading Massachusetts, 1965), Vol. III, formula (6.35).
- 280. Ref. 47, p. 1073.
- 281. W.H. Furry, Phys. Rev. **49**, 393 (1935).
- 282. E. Schrödinger, Proc. Cambridge Philos. Soc. **32**, 446 (1936).
- 283. P.H. Eberhard, Nuovo Cimento B **46**, 392 (1978).
- 284. G.C. Ghirardi, A. Rimini, and T. Weber, Lett. Nuovo Cimento **27**, 293 (1980).
- 285. A. Berthelot, Nuovo Cimento B **57**, 193 (1980).
- 286. P.J. Bussey, Phys. Lett. A **90**, 9 (1982).
- 287. T.F. Jordan, *ibid.* **94**, 264 (1983).
- 288. P.J. Bussey, *ibid.* **123**, 1 (1987).

Arthur Jabs

Sektion Physik der Universität München
Theoretische Physik
Theresienstr. 37
D-80333 München, Germany

Permanent address:
Department of Physics
Federal University of Paraíba
58.059-970 João Pessoa, Brazil

This figure "fig1-1.png" is available in "png" format from:

<http://arXiv.org/ps/quant-ph/9606017v1>

This figure "fig1-2.png" is available in "png" format from:

<http://arXiv.org/ps/quant-ph/9606017v1>

This figure "fig1-3.png" is available in "png" format from:

<http://arXiv.org/ps/quant-ph/9606017v1>